

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Takanori OKURA
Kakuji TORIGOE
Masahi KURIMOTO
- (ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF
INDUCING THE PRODUCTION OF INTERFERON- γ
- (iii) NUMBER OF SEQUENCES: 35
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: BROWDY AND NEIMARK
 - (B) STREET: 419 Seventh Street, N.W., Suite 300
 - (C) CITY: Washington
 - (D) STATE: D.C.
 - (E) COUNTRY: USA
 - (F) ZIP: 20004
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: Patent In Release #1.0, Version #1.30
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: JP 185,305/96
 - (B) FILING DATE: 27-JUN-1996
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: BROWDY, Roger L.
 - (B) REGISTRATION NUMBER: 25,618
 - (C) REFERENCE/DOCKET NUMBER: OKURA=1
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 202-628-5197
 - (B) TELEFAX: 202-737-3528

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 157 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Tyr	Phe	Gly	Lys	Leu	Glu	Ser	Lys	Leu	Ser	Val	Ile	Arg	Asn	Leu	Asn	1	5	10	15
Asp	Gln	Val	Leu	Phe	Ile	Asp	Gln	Gly	Asn	Arg	Pro	Leu	Phe	Glu	Asp	20	25	30	
Met	Thr	Asp	Ser	Asp	Cys	Arg	Asp	Asn	Ala	Pro	Arg	Thr	Ile	Phe	Ile	35	40	45	
Ile	Ser	Met	Tyr	Lys	Asp	Ser	Gln	Pro	Arg	Gly	Met	Ala	Val	Thr	Ile	50	55	60	
Ser	Val	Lys	Cys	Glu	Lys	Ile	Ser	Xaa	Leu	Ser	Cys	Glu	Asn	Lys	Ile	65	70	75	80
Ile	Ser	Phe	Lys	Glu	Met	Asn	Pro	Pro	Asp	Asn	Ile	Lys	Asp	Thr	Lys	85	90	95	
Ser	Asp	Ile	Ile	Phe	Phe	Gln	Arg	Ser	Val	Pro	Gly	His	Asp	Asn	Lys	100	105	110	
Met	Gln	Phe	Glu	Ser	Ser	Ser	Tyr	Glu	Gly	Tyr	Phe	Leu	Ala	Cys	Glu	115	120	125	

Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
 130 135 140
 Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp
 145 150 155

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1120 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: liver

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION: 1..177
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 178..285
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 286..756
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: 3'UTR
- (B) LOCATION: 757..1120
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

GCCTGGACAG TCAGCAAGSA ATTGTCTCCC AGTGCATTTT GCCCTCCTGG CTGCCAACTC   60
TGGCTGCTAA AGCGGCTGCC ACCTGCTGCA GTCTACACAG CTTGGGGAAG AGGAAAGGAA   120
CCTCAGACCT TCCAGATCGC TTCCTCTCGC AACAAACTAT TTGTGCGAGG AATAAAG      177
ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG      225
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met
-35 -30 -25
AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC      273
Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn
-20 -15 -10 -5
CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA      321
Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile
1 5 10
AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT      369
Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro
15 20 25
CTA TTT GAA GAT ATG ACT GAT TCT TCT GAC TGT AGA GAT AAT GCA CCC CGG      417
Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg
30 35 40
ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG      465
Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met
45 50 55 60
GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT      513
Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys
65 70 75
GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC      561
Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile
80 85 90
AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA      609

```

Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly
 95 100 105

CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT 657
 His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe
 110 115 120
 CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA 705
 Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys
 125 130 135 140
 GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA 753
 Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu
 145 150 155
 GAC TAGCTATTAA AATTTTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA 806
 Asp
 GCCCTTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAAGTCA GGTGTTCAAG ACCAGCCTGA 866
 CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG 926
 CATGCCCTCA ATCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG 986
 GTAGAGGTTG TGSTGAGCCG AGATTGCACC ATTGCGCTCT AGCCTGGGCA ACAACAGCAA 1046
 AACTCCATCT CAAAAAATAA AATAAATAAA TAAACAAATA AAAAATTCAT AATGTGAAAA 1106
 AAAAAAAAAA AAAA 1120

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..135
- (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA 47
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
 -5 1 5 10
 GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT 95
 Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn
 15 20 25
 CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G 135
 Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp
 30 35 40

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION: 1..134
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC	47
Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser	
40 45 50 55	
CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT	95
Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile	
60 65 70	
TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG	134
Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys	
80 85	

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 87 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: human
 (F) TISSUE TYPE: placenta

(ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION: 1..87
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GAATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG	50
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val	
-35 -30 -25	
GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G	87
Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: human
 (F) TISSUE TYPE: placenta

(ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION: 1..87
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CT GAA GAT GAT G
Ala Glu Asp Asp Glu
-10

12

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2167 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon + 3'UTR
- (B) LOCATION: 1..2167
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA	48
Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile	
85 90 95 100	
TTC TTT CAG AGA AGT GTC CCA GBA CAT GAT AAT AAG ATG CAA TTT GAA	96
Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu	
105 110 115	
TCT TCA TCA TAC GAA GBA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC	144
Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp	
120 125 130	
CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT	192
Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser	
135 140 145	
ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTTCA TGCCGGGCGC	246
Ile Met Phe Thr Val Gln Asn Glu Asp	
150 155	
AGTGGCTCAC GCCTGTAATC CCAGCCCTTT GGGAGGCTGA GGCGGGCAGA TCACCAGAGG	306
TCAGGTGTTT AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA	366
AAAATTAGCT GAGTGTAGTG ACCCATCCCG TCAATCCAG CTAATCAAGA GGCTGAGGCA	426
GGAGAATCAC TTGCACTCCG GAGGTGAGG TTGTGTTGAG CCGAGATTGC ACCATTGCGC	486
TCTAGCCTG GCAACAACAG CAAACTCCA TCTCAAAAAA TAAAATAAAT AAATAAACAA	546
ATAAAAAATT CATAATGTGA ACTGTCTGAA TTTTATGTT TAGAAAGATT ATGAGATTAT	606
TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCAAT CTTGAAAAAC ATCATTAAAG	666
AATGAATGAA CTTTCACAAA AACAACAAA CAGACTTTCC CTTATTTAAG TGAATAAAT	726
AAAATAAAT AAAATAATGT TTAATAAAT CATAGTTTGA AAACATTCTA CATTGTTAAT	786
TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGGTTAT TAGTCCTAAT	846
GACAAAAGAT ATTGATATTT GAACTTTCTA ATTTTAAAGA ATATCGTTAA ACCATCAATA	906
TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAT TTCTCCCAA GTCAGTATAT	966
TTTTAAATTT CATTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC	1026
CAACTGACAT TCATCCTAGG AATGCAAGA TGGTTTAATA TCCTAAAATC AATTAACATA	1086
ACATACTATA TTAATAAAGT ATCAAAACAG TATTCTCATC TTTTCTCTT TTTTCACAAT	1146
TCCTTGGTTA CACTATCATC TCAATAGATG CAGAAAAAGC ATTTGACAAA ATCCAATTCA	1206
TAATAAAAT TCTCAAACTT GAAAGAGAAC ATCATAAAGG CATCTATGAA AAACCTACAG	1266
CTAATATCAT ACTTAAAGAT GAAAAACTCA ATTATTTTAC CCTAAGATCA AGAATAATGC	1326
AAGCATGTCA GCTCTTGCAA CTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA	1386
CCATACAATA AATAAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTTT ATTTGCAGAC	1446
AACATGGTTT TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT	1506
TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAAATA CATTGAAGGC TGGGCTCAGT	1566
GGAGATGGCA TGTACCTTT GTCCAGCTA CTTGGGAGG TGAGGTAGGA GGATCACTTG	1626
AGGTGAGGAG TTTGAGCTA TAGTGCAATG TGATCTTGCC TGTGAATAGC CACTGCACTC	1686
GAGCCTAGGC AACAAAGTGA GACCCCTCT CCAAAAAAAA AAATGGTATA TTGGTATTTT	1746
TGTATATGAA CAATGAATGA TCTGAAAAACA AGAAAAATTC ATTACAGATG GTATTAAAAA	1806
AATAAAATAC AAATAAATTT AGCAAAATAA TTATAAAACT TGTACATCGA AAATTTCAAA	1866

GCACTCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA 1925
 TTG3AAAATT CATTCAATAT TGTTAAGATA ACAATTGTCC CCAAATTGAT GCATGCATTG 1985
 AATTTAGTCT TCATCAAAAT TCCAGCAGG3 TTTTTCGAGA AATTGACAAG CTGTACCCAA 2045
 AATGTATATG GAAATGAAAA GACCCAGAA3 AGCAAATAAT TTTTAAAAA CAAAGTTGGA 2105
 AAACCTTTTAC TTCCTAATTT TAAACTTTAC TATAAACCTA AAGTTATCAA GACCATTTAG 2165
 T 2167

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1334 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: intron
 - (B) LOCATION: 1..1334
 - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTATTTTTTT	TAATTCGCAA	ACATAGAAAT	GACTAGCTAC	TTCTTCCCAT	TCTGTTTTAC	60
TGCTTACATT	GTTCCGTGCT	A3TCCCAATC	CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
ATTTCACTTA	CAGGAAACTT	TATAAGGCAT	CCACGTTTTT	TAGTTGGGGT	AAAAAATTGG	180
ATACAATAAG	ACATTGCTAG	G3GTGATGCC	TCTCTGAGCC	TGCCTTTGAA	TCACCAATCC	240
CTTTATTGTG	ATTGCATTAA	CTGTTTAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
GTTACAGCTG	AAAATGCTGA	TAGTTTACCA	GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
TACTTG33AG	GCTCAG3CAG	GAGGATTGCT	TGAGGCCAGG	ACTTTGAGGC	TGTAGTACAC	420
TGTGATCGTA	CCTGTGAATA	G3CACTGCAC	TCCAGCCTGG	GTGATATACA	GACCTTGTCT	480
CTAAAATTAA	AAAAAAAAAA	AAAAAAAAAC	TTAG3AAAGG	AAATTGATCA	AGTCTACTGT	540
GCCTTCCAAA	ACATGAATTC	CAAAATATCAA	AGTTAG3CTG	AGTTGAAGCA	GTGAATGTGC	600
ATTCTTTAAA	AATACTGAAT	ACTTACCTTA	ACATATATTT	TAAATATTTT	ATTTAGCATT	660
TAAAAGTTAA	AAACAATCTT	TTAGAATTCA	TATCTTTAAA	ATACTCAAAA	AAGTTGCAGC	720
GTGTGTGTTG	TAATACACAT	TAAACT3TGG	G3TTGTTTGT	TTGTTTGAGA	TGCAGTTTCA	780
CTCTGTCAAC	CAGGCTGAAG	T3CAGT3CA3	T3CAGTGGTG	TGATCTCG3C	TCACTACAAC	840
CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	G3CTCAGTCT	CCCGAGTAG3	TGGGATTACA	900
GGCATGCACC	ACTTACACCC	G3CTAATTTT	T3TATTTTTA	GTAGAGCTGG	GGTTTCACCA	960
TGTTG3CCCA3	GCTG3TCTCA	AA3CCCTAAC	CTCAAGTGAT	CTG3CTGCCT	CAGCCTCCCA	1020
AACAACAAA	CAACCCCA	GTTTAATATG	TGTTACAACA	CACATGCTGC	AACTTTTATG	1080
AGTATTTTAA	T3ATATA3AT	TATAAAAG3T	T3TTTTTAAC	TTTTAAATGC	TG3GATTACA	1140
GGCATGA3CC	ACTGTGCCAG	G3CTGAACTG	T3TTTTTAAA	AATGTCTGAC	CAG3CTGTACA	1200
TA3TCTCTCTG	CAGACTGGCC	AA3TCTCAAA	GTGGGAACAG	GTGTATTAA3	GACTATCCTT	1260
TG3TTAAATT	TCC3CAAAATG	TTCTGT3TCA	AGAATTCTTC	TAACTAGAGT	TCTCATTTAT	1320
TATATTTATT	TCAG					1334

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4773 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..4773
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTAAGACTGA	GCCTTACTTT	GTTTTCAATC	ATGTTAATAT	AATCAATATA	ATTAGARATA	60
TAACATTATT	TCTAATGTTA	ATATAAGTAA	TGTAANTTAA	AAACTCAAAT	ATCCTCAGAC	120
CAACCTTTT	TCTAGAACAG	AAATAACAAG	AAACAGAGAA	CCATTAAAGT	GAATACTTAC	180
TAAAAATTAT	CAAACCTTTT	ACCTATTGTG	ATAATGATGG	TTTTTCTGAG	CCTGTACAG	240
GGGAAGAGGA	GATACAACAC	TTGTTTTATG	ACCTGATCT	CCTGAACAAT	CAGTCTTTAT	300
ATAAATAATA	ATGTAGAATA	CATATGTGAG	TTATACATTT	AAGAATAATA	TGTGACTTTC	360
CAGAAATGAT	TCTGCTATGA	AGAAATGAAG	TAATTATCCT	TCTATATTTT	TACACCTTTG	420
TAAATTATGA	TAATATTTTA	ATCCCTAGTT	GTTTTGTGTC	TGATCCTTAA	CCTAAATCTT	480
AGACACAAGT	TTCAGCTTCC	AJTTATGTGA	TGTTATTTTT	AATGTTAATC	TAATTGAATA	540
AAAGTTATGA	GATCAGCTGT	AAAAATTAATG	CTATAATTAT	CTTCAAGGCA	GATATAAGT	600
ATTTCTGTC	TCTACTTTTT	CTCTATTATT	CTCCATTATT	ATTCTCTATT	ATTTTTCTCT	660
ATTTCTCTCA	TTATTGTTAG	ATAAACCCACA	ATTAACTATA	GCTACAGACT	GAGCCAGTAA	720
GAGTAGCCAG	GGATGCTTAC	AAATTGSCAA	TGCTTCAGAG	GAGAATTCTA	TGTCATCAAG	780
ACTCTTTTTT	AATGGAATAT	TGCCAATAAA	TATGCTCTTT	CATGCCCAAC	CAGTCCCAAC	840
TGAAGAGACG	TTAGGATATG	ACCTTATGTG	AGGTACCAAG	GGGCAACTTG	GTAGGAGAAA	900
AAAAGCCACT	CTAAATATA	ATCCAAATAA	GAACAGTGCA	TATGCAACAG	ATACAGCCCC	960
CAGACAATTC	CCTCAGCTAT	CTCCCTCCAA	CCAGAGTGCC	ACCCCTTCAG	GTGACAAATTT	1020
GGAGTCCCAA	TTCTAGACCT	GACAGGCAGC	TTAGTTATCA	AAATAGCATA	AGAGCCCTGG	1080
GATGGAAGGG	TAGGGTGGAA	AGGGTTAAGC	ATGCTGTTAC	TGAACAACAT	AATTAGAAGG	1140
GAAGGAGATG	GCCAACTCA	AGCTATGTG	GATAGAGGAA	AACTCAGCTG	CAGAGGCAGA	1200
TTCAGAAACT	GGGATAAGTC	CGAACCTACA	GSTGATTTCT	TGTTGAGGGA	GACTGCTGAA	1260
ARTGTTTABA	AGATGGAAT	AATGCTTGGC	ACTTAGTAGG	AACTGGGCAA	ATCCATATTT	1320
GGGGAAGCCT	GAASTTTATT	CAATTTTCAAT	GGCCCTTTTA	AAATAAAAGA	ATGTGGCTGG	1380
GGGTGCTTGG	TCACACCTGT	AATCCGAGCA	CTTTGGGAGG	CGAGGGGGGG	CGGATCCACT	1440
GAAGTCAGGA	GTTCAAGACC	AGCCTGACCA	ACATGGAGAA	ACCCCATCTC	TACTAATAAT	1500
ACAAATTTAG	CTGGGCGTGG	TGGCATATGC	CTGTAATCCC	AGCTACTCGG	GAGGCTGAGG	1560
CAGGAGATTC	TTTTGAACCC	GGAAGGAGCA	GSTTGGGATG	AGCCTAGATC	GTGGCATTTG	1620
ACTCCAGCCT	GGGCAACAG	AGCAAAACTC	GSTCTCAAAA	AAAAAAGAAA	AAAGTGAALA	1680
TTAACCAAGG	GCATTAGCTT	AATAATTTAA	TACTGTTTTT	AGGTAGGGGG	GGGGGTGGGT	1740
GGAAGAGATC	TGTGTAAATG	AGGGAATGTG	ACTATTTAAAC	TTGATCAGCA	TCATAGCAAA	1800
TCTGCTTCTG	GAAGGAACCT	AATAAATATT	AGTTGGAGGG	GGGAGAGAGG	TGAGGGGTGG	1860
ACTAGGAGCA	GTTTTAGCCC	TGCTCTTTAA	TGCTTTTTCG	TGCACTAAT	AAGGATCTTA	1920
GCAATGCTTA	TAAAGTGGC	CTAGGTTCTA	GATAAATAAA	TACAACAGCC	CAGGACAGT	1980
GGCTCATCTC	TATAATCCCA	GCCTTTGGG	AGGGCAAGGC	GAGTGTCTTA	CTTGAGATCA	2040
GGAATTTAAG	ACCAGCTGG	CCAGATGGC	GATACTCTGT	CTCTACTAAA	AAAAATACAA	2100
AAATTATCTA	GGCATGGTGG	CATGCACTTG	TAATCCCAAC	TACTGCTGAG	CCTGAGGCGAG	2160
ARGAATGCTT	TGAAGCAAGG	AGGTGTAGGC	TGCAATGAGC	TGAGATGGCA	CCACTGCACT	2220
CCAGCTCTGG	CGACAGAAAT	AGACTTTGTC	TCAAAAGAAA	AAAAAGATAC	AACAGGATAC	2280
CCTTATCTCT	TCACCTTTCA	CTGTTGATTA	CTAGCTATAA	AGTCTATATA	AGTCTTTTGG	2340
TCAAGAACCT	TGACAAACAT	AAGAGGGATT	TGCTTTGAGA	GSTTACTGTC	AGATCTCTTT	2400
TCATATATAT	ACATATACAT	GTATATATGT	ATCTATATCC	AGGCTTGGGG	AGGCTTCCCT	2460
CAGACTTTCT	AGTGCACTTG	GGAGATGTTA	GSTCAATATC	AACTTTCCCT	GGATTCAGAT	2520
TCAACCCCTT	CTGATGTAAA	AAAAAAGAAA	AAAAAGAGAG	AAATCCCTTT	CCCTTTGAG	2580
CAGTCAATTT	TGAAGAGGTG	GGGCTTTCCA	AGTTGGGGGT	TCTCCAGGTT	CATTGGGATT	2640
GCTTTCAATAT	CCATTTGCTA	TGTACCTTCC	CTATGATGTC	TGGGAGTGGT	CAACATCAAA	2700
ACTAGGAAAG	CTAGTGCCCA	AGGATGTCTT	TAGCTCATTT	CTGAAATTTG	CAATAAGTGT	2760
GATTAAAGAG	ATTGCTGTGT	CTACCTATCC	ACACTGCTGG	TTTCAACTGT	AACTTTCTTT	2820
TTTTCTTTT	TCTTTTTTTT	CTTTTTTTTT	GAAAGTGAAT	CTGGCTCTGT	CGGCAAGGCT	2880
AGAGTGGAGT	GSCAGATCT	CAGCTCATTG	CAGGCTGTC	CTCCCGGGTT	CAGGCTATTC	2940
TGCTGCTTCA	CCCTCCCAAG	CAGCTGCGAC	TACAGGCTTC	TGCAAGCTAG	CCCACTTAAT	3000
TTTTGTATTT	TTTAGTAGAG	ACGGGGTTTT	ACGGGTGTTG	CGAGGATGGT	CTGATCTCTC	3060
TGAATTTTGT	ATCCGCGCGC	CTCAGCTCTC	CAAAATGATG	GGATTACAGG	CGTGAAGCAT	3120
CGCAGCCGGC	TCAACTGTAA	CTTTCTATAC	TGTTTTCATCT	TCCCTGTGAA	TGTTACTAGA	3180
GCTTTTGAAG	TTTTGGCTAT	GGATTATTTT	TGTTTATATC	ATTAGATTCT	AGATTATCTC	3240
CAAAATGATG	CCCAAGCTTT	AGGCTCTCTT	CTAAATCTGT	ATATTGTAGA	CAGCTGAGAA	3300
AGTGGGTGTC	AATAGGGGAA	CTAGTTTATA	CTTTTATCAA	CTTAGGAGCC	AGACTTGTGT	3360
ATAAGGAGCA	AAGGTCAAGA	GTTATGACTA	CTGATCTCAC	AACTGATTGA	GAAATTTGGG	3420
ATAACCCGCT	GACCTCTGCC	ATCCAGAGTC	TTTCAGGAT	CTTTGAAGGA	TGAAGAAATG	3480
CTATTTTAAAT	TTTGGAGGTT	TCTCATCTAG	TGCTTATGAT	CATGGGAATC	TGTGCTGCCA	3540
TGAGGCCAAA	AAAGTCCCA	AAACATCTAC	TGCTTCCAGG	ATTAACATGG	AAAGAACTTA	3600
GSTGCTGCCC	ACATGTTCTG	ATCCATCTCT	CAAAATAGAC	ATGCTGCACT	AACAAGAAAA	3660

GTGCAGGCGAG	CACTACCCAGT	TGGATAACCT	GCAAGATTAT	AGTTTCAAGT	AATCTAABCA	3720
TTTCTCACAA	GGCCCTATTG	TGTGACTGAA	ACATACAAGA	ATCTGCATTT	GGCCTTCTAA	3780
GGCAGGGGCC	AGCCAAAGAG	ACCATATTCA	GGACAGAAAT	TCAAGACTAC	TATGGAAGTG	3840
GAGTGTCTTG	CAGGGAAGAG	AGAATCAAGG	ACTGCCAACT	GAGGCAATAC	AGTAGGCTTA	3900
CACAGGAAGC	CAGGGCTAG	CCCTACAACA	ATTATTGGGT	CTATTCACTG	TAAGTTTTAA	3960
TTTCAGGCTC	CACTGAAAGA	GTAAGCTAAG	ATTCTTGGCA	CTTTCTGTCT	CTCTCACAGT	4020
TGGCTCAGAA	ATGAGAACTG	GTCAGGCCAG	GCATGGTGGC	TTACACCTGG	AATCCCAAGC	4080
CTTTGGGAGG	CCGAAGTGGG	AAGGTCACTT	GAGGCCAGGA	GTTCAGGACC	AGCTTAGGCA	4140
ATAAAGTGAG	ATACCCCTCT	ACCCCTTCTC	TACAAAAATA	AATTTTAAAA	ATTAGCCAAA	4200
TGTGCTGTGT	TATACTTACA	GTCCCACTA	CTCAGGAGGC	TGAGGCAGGG	GGATTGCTTG	4260
AGCCCAAGAA	TTCAAGGCTG	CAGTGAGCTA	TGATTTCAAC	ACTGCACTTC	TGGCTGGGCA	4320
ACAGAGGAG	ACCCTGTCTC	AAAGCAAAAA	GAAAAAGAAA	CTAGAACTAG	CCTAAGTTTG	4380
TGGGAGGAGG	TCATCATCGT	CTTTAGCCGT	GAATGGTTAT	TATAGAGGAC	AGAAATTGAC	4440
ATTAGCCCAA	AAAGCTTGTG	GTCTTTGTCT	GAACCTCTACT	TAATCTTGAG	CAAAATGTGA	4500
CACCACTCAA	TGGGAGAGGA	GAAGAAGTAA	CTGTTTGATG	TATAGGGGAA	AACTAAGGCA	4560
CTGGAAGTGA	ATATGCATCC	CATGACAGGG	AGAATAGGAG	ATTGCGAGTT	AAGAAGGAGA	4620
GGAGGTCACT	ACTGCTGTTC	AGAGATTTT	TTTATGTAAC	TCTTGAGAAG	CAAACTACT	4680
TTTGTCTCT	TTGTAATAT	ACTTCAAAAC	AACTTCATA	TATTCAAATT	GTTTCATGTC	4740
TGAAATAATT	AGGTAATGTT	TTTTTCTCTA	TAG			4773

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8835 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 1..8835
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GTAAGAAATA	TCATTCTCT	TTATTTGGAA	AGTCAGCCAT	GGCAATTAGA	GGTAAATAAG	60
CTAGAAAGCA	ATTGAGAGGA	ATATAAACCA	TCTAGCATCA	CTACGATGAG	CAGTCAGTAT	120
CAACATAAGA	AATATAGGCA	AAGTCAGAGT	AGAATTTTTT	TCTTTTATCA	GATATGGGAG	180
AGTATCACTT	TAGAGGAGAG	GTTCTCAAAC	TTTTTGCTCT	CATGTTCCCT	TTACACTAAG	240
CACATCACAT	GTTAGCATAA	GTAACATTTT	TAATTAAAAA	TAACTATGTA	CTTTTTTAA	300
AAACAAAAAA	AGCATAAGGA	GTGACACTTT	TTTATTTTTA	CAAGTGTTTT	AACTGGTTTA	360
ATAGAAGGCA	TATAGATCTG	CTGGATTCTC	ATCTGCTTTG	CATTGAGACT	ACTGCAATAT	420
TGCACAGAA	GCAGCTCTCT	GTAACCTCTG	TTGTACACTC	ATGAGAGAAT	GGGTGAAAAA	480
GACAAATTAC	GTCTTAGAAT	TATTAGAAAT	AGCTTTCACT	TTAGGAACTC	CCTGAGAATT	540
GTTGCTTTAG	AGTGGTAAGA	TAAATAAGCT	TCTCTTTAAA	CGGAATCTCA	AGACAGAAAT	600
ASTTACATTA	AAAGCAAAAC	AAAAATTTGC	CCATGGTTAG	TCATCTTGTT	AAATCTGCCA	660
CACTCTTGG	CTGGGCTACA	ATTGGATAAT	ATAGCATTC	CCGAGATAAT	TTTCTCTCAC	720
AATTAAGGAA	AGGGCTGAAT	AAATATCTCT	GTTTGAAGTT	GAATAACAAA	AATTAGGACC	780
CCCTAAATTT	TAGGCTCTCT	GAAATTCGTC	TTTTTGCCTA	TATTCAGCTA	CTTTACGTTT	840
TATTAAATCT	TCTTTTCAGG	CAGGTGCACT	AGCTCATGCC	TAGAATCTCA	GTCAGGCTCT	900
AGCCCAAGAA	TTTGAAGCCA	GCCAGGGCAA	CACAGTCTCT	ACAAAAAAT	AAAAAATTAC	960
CTGGCTGTGT	TGGTGCATGC	CTGTAAGAA	ACTCAGGATG	CTGAGGACTG	CTTGAGGCTCA	1020
GGATAGCCAA	ATCTGTGGTG	AGTTCAAGCA	CTAAACAGAG	CSAGACTTTT	TCAAAAAAAC	1080
AAACAAAAAA	ACAAACAAAC	TTCTTTCAAA	ATACTTTTTT	ATCTGCAATG	TTTTCTTATT	1140
GCCTGTGAGA	TAAATTTTAC	TCTTTTCACT	GATTTCCAAA	GCCCTCCATA	ATCTAATCCG	1200
ACTTTACCTT	GTGTTCACTG	CAAAATAGCA	GGACTGTTTC	ACTACAAATC	AAAAATCACA	1260
GTTTGGGGTG	AGTGGCTCAC	TCTCTAATC	CCAACACTTT	GGAAGGCCAA	GTCAGGCTCA	1320
TTGCTTCAGC	TCAGGAGTTC	AAGACCAGCT	TGGGCAACAT	GGCAAAAAAC	CTGTCTCTCC	1380
AAAACATACA	AAAATTAGCC	AGATGTGGCA	GATGTGCCCT	GTAATCCCAA	CTACTCAAAA	1440
GGCTAAGGCA	AGAGGATCAC	TTGAGCCGAG	GGGTCAAGG	CTACAGTGA	CCATGTTTAC	1500
TGTGTCACTG	CACTCCAGCC	TGGGTGATAG	AGCAAGACCA	TGTCTCAAAA	AAAAAATAAA	1560

GAAAAAAGAAA	GAAAAAAGAAA	TGCTCTATT	CAGTTTACCT	CCACACAAAC	ATTGTTTTTA	1620
TTATACATA	AATGCTGGTC	CATTGCTTC	TCTATCTATT	CAATCTTTA	AGCATTCTTT	1680
GAGATTCAA	TCAATTCTCC	TTTTAAACT	AGGCTATTTA	AACTACATCA	GTTCCTTTTT	1740
GATTTTCTT	CTTTGAGTCT	ATAGACTCAA	AAACAAAAAC	TTAAAAACTT	ATTTTTTTAA	1800
TTTTCTGCTA	CTCTCACTTC	TTCAACACTC	ACATACACGC	ATTCTAATA	AGATGGCAGA	1860
ATGTTCAAAG	ATAAAATGAT	TTATAGAACT	GAAAAAGTTAG	GTTTTGATCT	TGTTGCTCTC	1920
AAGATGACTA	CCTACCTGAT	CTCAGGTAAT	TAATTATGTA	GCATGCTCCC	TCATTTTCATC	1980
CCATACTTAT	TCAACAGGAT	TGAAATTCOA	CAGCAAGGAT	AAACATAATC	ATAGTTGCTT	2040
TTCAAATTTCA	AGGCATTTTA	ACTTTTAATC	TAGTAGTATG	TTTGTGTTTG	TTGTTGTTGT	2100
TTGAATGGA	GCCTGCTGT	GTCACTCAAG	CTGGAGTGCA	GTGCTACGAA	CTCGCTCTAC	2160
TGCAACCTCT	GCCTCATGCG	TTCAATCAAT	TATTCTGCTT	CAGTGTCCCA	AGTAAGCTGCT	2220
ACTACAAGGC	ACATGCCACC	ATGCTGCTCT	AATTTTTGTA	TTTTTAAGTA	AAACAGGCTCT	2280
TCACCATGTT	GGCCAGGCTG	GTCTGCAACT	CCTGACCTCA	AGTGATCCAC	CCGCTCTGCT	2340
CTCCCAAGT	GCTGGGATTA	CAGGCATAAG	CCACCGTGCT	CAGCTTAATA	GTATGTTTTT	2400
AAACTCTTAG	TGGCTTAACA	ATGCTGCTTG	TATAATAAAT	ATGCCATAAA	TATTTACTGT	2460
CTTAGAATTA	TGAAGAGGTG	GTTACTAGGC	CGTTTTGCCAC	ATATCAATG	TTCTCTCTCT	2520
ACAGCTTTAA	TTAGAGTCTA	GAATTGCTAG	TTGGTAGAGC	TGGAACAGAC	CTTAAGAGAT	2580
GACTAGCCAA	CTTCTCTGTC	CAAAATGAGG	AACTGAGAC	CTTAAATTA	AGTGACTTGC	2640
CCCAGACAAA	ACTGGAACCT	ATGCTGCTCA	ATTTCCATCA	TGAAATTTCT	CCATTCATCA	2700
GCCTCTGCT	AGTTGCTCAA	GTATGCTATA	ACTAAATTTT	TATGCTCTGT	TTAAAGACAA	2760
AATTGCTCACT	GCTTACTCCT	GGAAGGCTCT	TTCTGAGGTC	GTTTATAACT	CTTAAAGAAA	2820
AAAAAGTCAG	TAGTCTGAGA	ATTTTAGAGC	AAATAGTCAA	AGCATTTTTA	TCCAATGGAT	2880
CTATAATTTT	CATAGATTAG	AGTTAAATCA	AAGAACACGC	GATGAGAAAG	GAAGAGGAAA	2940
ATTGAGGAGA	GGAGGAATG	GGATGAGAAC	ACACTACTTG	TAATCAGTCA	TAGATGACT	3000
GAGAACTAAC	AAGAGGAAT	GTAAAGAAAT	AAGAATGAG	AATTCAAAT	CAAGCATGA	3060
AATAAAGAGA	AACCTAGAG	GAAAAATGGA	GAGACATTA	GAAAAATTAT	TCTATTTTTA	3120
AAATTTCTGT	TTCAAGGCTT	CTCTCTGTTT	TTCTCTCTTC	TCATTGGTTT	TCAGGCTGAG	3180
GGAATTTTA	AGATGGAAAA	AATATATATA	TTCTACACAT	CCCTTTCTAC	GCTGTTGCTA	3240
TGGCAACAG	GTTTATCATA	GCAAACTTTT	ATTCATACAA	CATTTATTGA	GTTCTTACTG	3300
TGTGGTAGG	TCTTTCCAG	TGTTGAAAT	TCAGGGGAAA	AAAGACAACT	CATTGCTCTA	3360
AAACTCAGAT	GAAAGCTGAA	CAGACCTATT	TTTAATCATA	GTAATCTCAA	TTTAGGCTAG	3420
TAGAGCTAT	TTAAGAGGCA	TGAACAGGTC	TGAAGGAGGT	AGGACTCTGA	GGAGAGATA	3480
GTTAGCTAG	AATGAAGAG	CAGAGAGATT	TTCTAGAGG	AACTATTAA	GCTGGGASTT	3540
ACGGGATGAA	AGATGAGGCA	GCGTTTCTAG	GCAAAAAAAA	AAAAAAGGCA	GGGAGGGG	3600
AAGTTCTGGC	CTGGCTAGGA	GAATAGCTGT	GCAACAAATG	GAGGAGAGTC	TGGAAGCTAG	3660
AAACCCAGT	AGAAGAGTAT	TAAATAGAA	GATGCCAGGG	GTAATGAGGG	CTTGATTTAA	3720
AACAGTGTG	TTGAGAGTGG	AGAGGAGATA	CCAAATTTCT	GAGACATTTT	TGAGTTAGAA	3780
CCTACAGTAT	TTATCAGACA	AAGGAAAGAT	TAGACAAAGG	AGTTAAGAA	GACTCCCGAG	3840
TTTCASTTTT	GGGCAAGTAA	CTAGGACATG	TTTTGMAAAG	TAAATGTTAT	GATCTCTTAC	3900
CATTGCAACT	ATGTTAGTGG	AGCTAAATTA	AAATTTGTAC	ATGTATATAA	CTCTCCCGCC	3960
ACCAACAGTA	ACTACTTCCC	TAAATCTCTA	CTTTGTAGCC	AGACTTCTTA	AAAGAAATAG	4020
TTTAGTTCAC	TGTCTTTTACT	TTTCCCTCTC	CATTCTGTCT	TAGATATTGG	TCCACCTAC	4080
ATCTGCTGCG	TCCACTTTTAC	CCAACTCTTT	CTAAGGTTGC	CCAAACTTTC	CTAATTTCTA	4140
AATTCATATT	ACAAGTTTAA	GCTTATATGT	AAATTAGGAG	CTCTACAGTT	TGATTTGAG	4200
CAGCCCTCTC	TGAAGCCCTT	TCTCTCTGCA	CTCTCTGAG	ACATCTCAGA	TTTACAAAG	4260
TGAAGTAAT	ATTTTACACT	TGAAGCTGAT	TTTCTGTTCT	CTTTCTTCT	GAATGAGGTA	4320
ACCACTCAAC	AAATTGCCCC	AGGCAAAAG	TAGAAAGTCA	TCTTCAGTTC	CTCCTCTCTC	4380
TGTTTGAAGC	ACAAGAGATC	AGCTGAGAAA	TCCCGCTGTT	TAGTATCTCT	TGAATTCATT	4440
ACCTTAATTT	ATAGCTCTCAT	CAACTCTTAA	TGTTTAAAT	TACTTCAGTA	GTTGTTCTCT	4500
GACTCTGCTC	CAATCTGCTT	CAATGAGGTC	CATCTCTTTG	TCTCTGCTGG	TGGTGGTGGT	4560
GTGACAGAG	TTTTCTCTTT	GCTTCCAGG	CTGAAGTGA	GTGAGGCACT	TCATTCAGAA	4620
CACAGCTCTC	TGGTTTAAAG	CAGTTTCACT	TCTGAGTAG	CTGGGAGTAC	AGGTATCTGC	4680
CATACAGGCT	AGCTAATTTT	GCTTTTTCAG	TAGAGACAGG	GTTTCACTAT	GTTTCTCAGG	4740
CTGCTCTCAA	ACTCTGAGCC	TCAAGCAATC	CAGCTCACTC	AGCTCTCCAA	AGTCTCTGGA	4800
TTACAGGCTAT	GAGCTCACTG	AGAGGAGGCA	GATCCATTGT	TTATGTCTCT	TCTAGAGTGA	4860
GTTTITTAATA	CACAAATTTG	ACCATATCTT	TCTCTCAATT	AACTCAATAT	TTTTTTTTTT	4920
AGGAAAAAAC	AGTTCAAACT	CTTTAGTCTG	CTTACACAA	GCTTTTGTAG	TCTGACTCTT	4980
CTTTTCAAGC	TTTCTACAAA	GTATCTGCTA	AGTTACATTT	TATCTGAATT	GAATTAAGCA	5040
AGCTATATAA	AATTATAGTT	TATATGCTCA	AAATGGAAT	AATGTTAACT	CTTCCAAATA	5100
GTTTATCTAG	AATGACATAA	TTTCAAAAGT	GTCAAGTCAA	ATGAGTTATA	AACTCTTAAC	5160
ACTATTGCTA	CATGCAAGTG	TCTCTTATAC	TTCTGAGAA	TATCTCTTCT	CATGCTCATTA	5220
TTATCTAAAT	TAGACTTTAA	ATAACTCAGA	AGTTCTTCTG	ACATACAGCT	TATTATCTGT	5280
CTTTTAAAC	ATAATTTTAA	ATAATTTTAT	ATATGATAAT	GTTATCCAA	TGCTAAGGGA	5340
TGTATTGTTA	CTGCTGTGCA	AAAAAAGAAA	AAAAAAGAAC	TCCAAATAAA	TATGTTGAAA	5400
CCAGGTTTAT	ATGCAAGAAA	ACAATATTAA	AAAGGCAAA	GTACCACCAT	AATAGGCTGT	5460
GTGAGAGGCT	CAGGCTACAA	AACACTAGTA	ATAATGCTGA	GAAAGTTGAA	AAAAAGAGAA	5520
AAGCAACAAT	ATGCTTTTGT	TGTTGTAGCT	TTATGTAATC	CAAGAAATAT	TCTCTCAAAA	5580

CTTTTAC3TT	TTTTCCAAAG	AAAASTTAAC	TTTGGCTGGG	CGCAGTGGCT	CTTGCTGTGA	5640
GTCCCCAGCT	TTGGGAG3CC	AAGGCGG3CA	GATCACCTGA	GGTCAGGA3T	TTGAGACCAG	5700
CCTGACCAAA	AATG3AGAAA	CCCGCCCCCC	TCACTACTAA	AAGAATACAA	AATTAGGCCG	5760
GGCACASTGG	CTTACCCCTG	T3ATCCCA3G	ACTTTGGGAG	GGCGAA3CAG	GAAGATCACC	5820
TGAG3TCA33	AGTTCCGA3C	CAGCTAT33A	GAAACCCGTC	TCTACTAAAA	ATACAAAATT	5880
AGCGGG3CT	GGT3GT3CAT	GACT3TAATC	CCAGCTACTC	AG3AG3CTAA	GGCAGAGAAT	5940
CACTTGAACC	CAGGCASTGG	AS3TTSCAST	GAG3CGAGAT	CGTGCCATTG	CACTCCAGCC	6000
TGG3CAACAA	GAGCGAAAAT	CTGTATCCAA	AAAACAAAAG	AAAAGAAAAG	GTAACTTTGA	6060
ACTATGTGAG	ATCTTTAGAA	ATGCATTCTT	TCTGTAAAAT	GTGACTACAT	TTGCCCTTAT	6120
TATGGTAAAA	ATGTTGAB3C	CTCAACAAC	CCATATTTTC	TCGGTCTCCC	CGTGCCTAG	6180
CCTTTGTTC	CATTGTCTCT	TCTTGT3GA	AGCTCTTCCT	CTGGCCTTGA	AAATGCCTGC	6240
TTCTCTTTCA	AG3TAGCACA	GTCATCACTT	TCTGTGGTAA	CCTTCTCCAG	CA3CATCAAA	6300
CA3AAAGAAT	GAATCTCTTG	TAAATTCAGC	TCTTACGTCA	TTCATTACAT	TATTTTGTAA	6360
CTCTTTATAG	ATTCTTCTCT	CCCACTAGAC	TCTGAGTCAC	TGGAGAGTAG	GAGCCAACTC	6420
TCATTCATCT	GTG3TTTGGT	CAGCTACTGG	CCACATTCCCT	GATGCATAGT	TAATGCTCAA	6480
ACCTTAACCTG	GTGAATCAGC	TCAAAATATTG	TCCTTCTCTA	AATCCATTCA	CTCATTGACT	6540
AACTATGTAC	TCAAAATAGT	AAACACCACT	AATTTAATCC	AATTCCTGCC	CATACTGCTT	6600
G3TACATTTT	AGGTGAATTA	GTTTGATAAA	TATGTGTGTA	TTACATAATA	TTAAAGTATG	6660
TACAGAAAGAT	CATGCTAATC	ATAATTCACA	ACTGATAACT	AATCAAACAT	AAATGCTCTC	6720
AG3TTAACAA	ATGTCTGCCCT	TCTCAGTTAA	TGCAGTCATT	AACAAACACC	TTCTGATGCT	6780
GATAATAGGG	CCTTGTTCAG	CAATGAAGCC	ATAAAGGTGA	ATAAAGAACA	TGCCCTCGTG	6840
GAGCTCACAG	CCTAGTCACT	ATTGTTCTGA	TTTTTAATAT	TAATGTTGGT	TTGGGTTTTG	6900
GT3AAAAATG	TTTAGACTTA	TCTTAGTGAT	CITTTTCATCC	TTTGCTATAT	TATTTTTCTC	6960
TAAGAGTCTT	CCTTATCCCC	TCTTTTAAAA	AAC TAGGTGA	TAATTTCTAA	TTGTAAATTT	7020
AAATATTATA	AATAGCTTAT	AAAATTTAAT	ATTTATAATA	TTTAAATGTT	TGATAAATAT	7080
TTAAATTTTA	TAATATTTAA	ATGTTTATTT	AAATTCATTT	GTACATCAGT	TTTTATTTTA	7140
TTTAAAT3TG	TTGGCCAG3C	ATG3T33CTG	ACACCTATAA	TCCCAGAACT	TTGAGAGGCC	7200
AA3TCAG3CA	AACCATTTGA	GCTCAG3AGT	TTGAGACCAC	CCTGGGCAAC	GTGGTGAAC	7260
CCTGTCT3TA	CCAAACATAT	GAAACCTTAT	CTG33TGTGG	TGGCACGCAT	CTGTGGTCCC	7320
AGATG33AGT	CCCAG3CTAA	GAT333AGAA	TCC3TTGAAC	CCAGGTGAGA	GGGGTGGGGT	7380
GGATGTT3CA	GTGAGCTGAG	ATC3T33CAC	TGCATCCCAA	CCTGGGTGAC	AGAGTGAGAC	7440
TCCATCTCAA	AAAAA3333A	T3TTATCTAA	ATAA3ATAAA	TTTAATAACT	GTTCCGCACTT	7500
AGATGAGCAT	AG3GAAT3AA	A3CTA3ATAA	A3CTATCAAA	TAAGGCCTGG	GTACAGTGAC	7560
TCATGCCCT3T	AATCTCAAGC	A3CTTT333AG	GGCAAAATTA	TACAAAGTTA	GTTGTATAAC	7620
ACCAACTAAC	A3CTATTTTG	G3GTTA33CT	AATTCAGATT	AATTTTTTTT	AAACTGAGTT	7680
TTAAATTCCT	GCTTACTCTA	CCATACATGC	TA3GCCTCAT	ATTATGCTAG	AAAAATTTTG	7740
AGCACAGATT	TATGAATACT	CTCCT3CATA	CCATTTAATT	TTTAAACAAA	TTTTAATGCA	7800
GTATATAT3T	GCCTTTTTAC	CAACACATTA	AATAATAAGA	TCTACTGTGA	GGACTAAATT	7860
TCTGTAAATT	CMAAGTAGTA	ATGAGTTTAA	ACCATGTCTC	AAGATCTCTG	CAATAACTGT	7920
AGCACACAG	AAAATA33TA	TTTCTATTAA	TGACAGAGTC	ACAAGTACTA	CTAATAATAC	7980
T3T3GTTT3T	TT3CT3CAAC	TATCTAT3GG	AG3AATGCTA	AATTTTCAGAG	GTTGGTGAAA	8040
ATACATGT3T	ATTTTTTTTC	CCAT3CAAGT	TCACAGATTT	CTCACACTGA	GAACCTCTAT	8100
TCCATAACRA	AATTTCT33A	G3CT3CACAC	CGTATTGGAA	GAAGGGCAGA	AAGGAAAAGC	8160
AAATGGAA33	ATTTAAATTT	TTTTCAAAATC	CTGTATCCCT	TGATTTTACA	GCAAGATTGT	8220
ATTTAT3TAT	TACTT3T3TT	AAAAATATAG	TATATTCGAG	ACTCCAGATC	AAAAATCACC	8280
GCAGCTCA33	GAGAAA3AG3	33CACTAAAT	G3CAGAGCCC	TTCAGCCTTC	TCCCACCCCTG	8340
CCTGTACCCCT	CAGAT33AA3	CA3TTTTTTA	TCATTGTTTC	ACCTTTAGCA	TTTTGACAAT	8400
GAAGTCACAA	ACCTTCAB3C	TCTCAACCAT	AG3AACCCAC	TGGTTGTAAG	AGAAGGATGA	8460
AGCCAGT3CT	TCCTAAAG33	CAC3ATTAGA	T3T3TTTATG	GCATCCTCAG	GTGAAACTAT	8520
ATTTATATTG	ACAATATATT	TATATTTCTC	AG3AATAACT	AG3AATAATGA	TTGAGTT3AG	8580
TACTAG33CA	TTTATCTATC	CTTTATAATA	TT3TTTAAAT3	AG3AAATGCT	TTCTATCTTC	8640
CMAATATT3G	AT3ATTT3TA	AGAGAACACT	TAAACATG33	TATTCATAAG	CTGAAACTTC	8700
T33CATTTAT	TGAAT3TCAA	GATT3TT3AT	CAGTATACTA	G3TGATTAAC	T3ACCACTGA	8760
ACTTGAA3GT	AGTATAAG3T	AGTAGTAAAA	G3TACAATCA	TTGTCTCTTA	ACAGATG3CT	8820
CTTT3CTTTT	ATTAG					8835

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 1..1371
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTAAGGCTAA	TGCCATAGAA	CAAATACCAG	GTTCA3ATAA	ATCTATTCAA	TTAGAAAAGA	60
TGTTGTGAGG	TGAACTATTA	AGTGACTCTT	TGTGTCAACCA	AATTTCACTG	TAATATTAAT	120
GGCTCTTAAA	AAAATAGTGG	ACCTCTAGAA	ATTAACCACA	ACATGTCCAA	GGTCTCAGCA	180
CCTTGTCACA	CCACGTGTCC	TGGCACTTTA	ATCAGCAGTA	GCTCACTCTC	CAGTTGGCAG	240
TAAGTGACAA	TCATGAAAAT	CCCAGTTTTT	ATGGGAAAAT	CCCAGTTTTT	ATTGGATTTC	300
CATGGGAAAA	ATCCCAGTAC	AAAACCTGGT	GCATTCAAGG	AATACAATTT	CCCAAAGCAA	360
ATTGGCAAAAT	TATGTAAGAG	ATTCTCTAAA	TTTAGAGTTC	CGTGAATTAC	ACCATTTTAT	420
GTAATATATG	TTGACAAGTA	AAAATTGATT	CTTTTTTTTT	TTTTCTGTTG	CCCAG3CTGG	480
AGTGCAGTGG	CACAATCTCT	GCTCACTGCA	ACCTCCACCT	CCTGGGTTCA	AGCAATTCTC	540
CTGCCTCAGC	CTTCTGAGTA	GCTG33ACTA	CAGGTGCATC	CCGCCATGCC	TGGCTAATTT	600
TTGGGTATTT	TTACTAGAGA	CAGG3TTTTG	GCATGTTGTC	CAGGCTGGTC	TTGGACTCCT	660
GATCTCAGAT	GATCCTCCTG	GCTCG3GCTC	CCAAAGTGCT	GGGATTACAG	GCATGAACCA	720
CCACACATGG	CCTAAAAATT	GATTCTTATG	ATTAATCTCC	TGTGAACAAT	TTGGCTTCAT	780
TTGAAAGTTT	GCCTTCATTT	GAAACCTTCA	TTTAAAAGCC	TGAGCAACAA	AGTGAGACCC	840
CATCTCTACA	AAAACCTGCA	AAATATCCTG	TGGACACCTC	CTACCTTCTG	TGGAG3CTGA	900
AGCAGGAGGA	TCACTTGAGC	CTAG3AATTT	GA3CCTGCAG	TGAGCTATGA	TCCCACCCCT	960
ACACTCCAGC	CTGCATGACA	GTAGACCCCTG	ACACACACAC	ACAAAAAATA	ACCTTCATAA	1020
AAAATTATTA	GTTGACTTTT	CTTA33TGAC	TTTCCGTTTA	AGCAATAAAT	TTAAAAGTAA	1080
AATCTCTAAT	TTTAGAAAAT	TTATTTTTAG	TTACATATTG	AAATTTTTTA	ACCCTAGGTT	1140
TAAGTTTTAT	GTCTAAATTA	CCTG3AACA	CACTAAGTCT	GATAAGCTTC	ATTTTATGGG	1200
CCTTTTGGAT	GATTATATAA	TATTCTGATG	AAAGCCAAGA	CAGACCCTTA	AACCATAAAA	1260
ATAGGAGTTC	GAGAAAGAGG	AGTA3CAAAA	GTAAAAGCTA	GAATGAGATT	GAATTCTGAG	1320
TCGAAATACA	AAATTTTACA	TATTCTGTTT	CTCTCTTTTT	CCCCCTCTTA	G	1371

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 1..3383
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTAAAGTAGA	AATGAATTTA	TTTTTCTTTG	CAAACTAAGT	ATCTGCTTGA	GACACATCTA	60
TCTCAACATT	GTCAGCTGAG	GAATAAAAAA	AATGGTTCTC	ATGCTACCAA	TCTGCTTCTA	120
AAGAAATGTG	GACTCAGTAG	CACAGCTTTG	GAATGAAGAT	GATCATAAGA	GATACAAAGA	180
AGAACTCTTA	GCAAAAGATG	CTTCTCTATG	CCTTAAAAAA	TTCTCCAGCT	CTTAGAATCT	240
ACAAAATAGA	CTTTGCTTGT	TTTATTGCTC	CTAAGATTAG	CATGAAGCCA	TGGATTCTGT	300
TGTAG3GGGA	GCGTTGCATA	GGAAAAAGGG	ATTGAAGCAT	TAGAATTGTC	CAAAATCAAT	360
AACACCTCCT	CTCAGAAATG	CTTTG3GAAG	AAGCCTGGAA	GGTTCCGGGT	TGGTGGTGGG	420
GTGGG3CAGA	AAATTCTGGA	AGTAGAG3AG	ATAGGAATGG	GTG3GGCAAG	AAGACCACAT	480
TCAGA3GCCA	AAAGCTGAAA	GAAACCATGG	CATTTATGAT	GAATTCAGGG	TAATTCAGAA	540
TGGAAGTAGA	GTAGGAGTAG	GAGACTGGTG	AGAGGAGCTA	GAGTGATAAA	CAGG3TGTAG	600
AGCAAGACGT	TCTCTCACCC	CAAGATGTGA	AATTTGGACT	TTATCTTGGG	GATAATAGGG	660
TTAATTAAGC	ACAATATGTA	TTAGCTAGGG	TAAAGATTAG	TTTGTGTGTA	CAAAGACATC	720

CAAAGATACA	GTAGCTGAAT	AAGATAGAGA	ATTTTTCTCT	CAAAGAAAGT	CTAAGTAGGC	780
AGCTCAGAAG	TAGTATGGCT	GGAAGCAACC	TGATGATATT	GGGATCCCCA	ACCTTCTTCA	840
GTCTTGATCC	CATCATCCCC	TAGTTGTTGA	TCTCACTCAC	ATAGTTGAAA	ATCATCATAC	900
TTCTTGSGTT	CATATCCCAG	TTATCAAGAA	AGGGTCAAGA	GAAGTCAGGC	TCATTCCCTT	960
CAAAGACTCT	AATTGGAAAT	TAAACACATC	AATCCCCCTC	ATATTCCATT	GACTAGAATT	1020
TAATCACATG	GCCACACCAA	GTGCAAGGAA	ATCTGGAAAA	TATAATCTTT	ATTCCAGSTA	1080
GCCATATGAC	TCTTTAAAAAT	TCAGAAATAA	TATATTTTTA	AAATATCATT	CTGCTTTTGG	1140
TATAAAGAAT	TGATGGTGTG	GGGTGAGGAG	GCCAAAATTA	AGGGTTGAGA	GCCATTATT	1200
TTAGTTATTA	CAA3AAATGA	TGGTGTCTAT	AATTAAGSTA	GACATAGGGG	AGTGTCTGAT	1260
A33AGCTGTG	AATG3ATTTT	AGAAACACTT	GAGAGAATCA	ATAGGACATG	ATTTAGGGTT	1320
G3ATTT3GAA	AGGAGAA3AA	AGTAGAAAAG	ATGATGCCTA	CATTTTTTCAC	TTAGGCAATT	1380
TGTACCATTC	A3T3AAATAG	GGAACACAGG	AG3AA3AG3A	G3TTTTG3TG	TATACAAAGA	1440
G3A33AT33A	TGA3G3ATTT	CGTTTTGGAT	CTGAGATGTC	T3TG3AACGT	CCTAGTGGAG	1500
ATGTCCACAA	ACTCTTCTAC	ATGTG3TTCT	G3TTTCAG3A	CACAGATTTG	GGCTGGAGAT	1560
AGAGATATTG	TAG3CTTATA	CATAGAAATG	GCATTTTGAAT	CTATAGAGAT	AAAAAGACAC	1620
ATCA3AG3AA	AT3TG3TAAAG	TGAGAGAGGA	AAAGCCAAGT	ACTGTGCTGG	GGGGAATACC	1680
TACATTTAAA	GGATG3CASTA	GAAAGAAG3T	AATAAACAAC	A3AGAGCAGA	CTAACCAAAA	1740
G333AGAA3A	AAAACCAAGA	GAATTCACCC	GACTCCACAG	A3AGCATTTT	AAGATTGAGG	1800
GGATAG3TGT	T3TGTTGAAAT	TTTGCA3CCT	TGAGAATCAA	G3GCCAGAAC	ACAGCTTTTA	1860
GATTTAG3AA	CAAG3AGTTT	GGTGATCTCA	GT3AAAGCAG	CTT3ATGGTG	AAATGGAGGC	1920
AGAG3CA3AT	T3CAAT3AGT	GAAACAGTGA	AT3G3AAGTG	AAGAAATGAT	ACAGATAATT	1980
CTTGCTAAAA	GCTTG3CTGT	TAAAAGGAGG	AGAGAAACAA	GACTAGCTGC	AAAGTGAGAT	2040
TG3GTTGATG	GAGCAGTTTT	AAATCTCAAA	ATAAG3AGCT	TTGTGCTTTT	TTGATTATGA	2100
AAATAAT3TG	TTAATTGTAA	CTAATTGAGG	CAATGAAAAA	AGATAATAAT	ATGAAAGATA	2160
AAAATATAAA	AACCACCCAG	AAATAATGAT	AGCTACCATT	TTGATACAAT	ATTTCTACAC	2220
TCCTTTCTAT	GTATATATAC	AGACACAGAA	AT3CTTATAT	TTTTATTAAA	AGGGATTGTA	2280
CTATACCTAA	GCTGCTTTTT	CTAGTTAGTG	ATATATATGG	ACATCTCTCC	ATGSCAACGA	2340
GTAATTGCAG	TTATATTAAAG	TTCATGATAT	TTCA3AATAA	G33CATATCT	TTGCCCTTTT	2400
TATTTAATCA	ATTCTTAATT	G3TGAATGTT	TGTTTTCCAGT	TTGTTGTTGT	TATTAACAAT	2460
GTTCCCATAA	GCATTCCTGT	ACACCAATGT	TCACACATTT	GTCTGATTTT	TTCTTCAGGA	2520
TAAAACCCAG	GAGGTAGAAAT	TGCTGG3TTG	ATA3AAGAGA	AAG3ATGATT	GCCAAATTAA	2580
AGCTTCASTA	GAGGGTACAT	GCCGAG3ACA	AAT3GGATCA	GCCCTAGATA	CCAGAAATGG	2640
CACTTTCTCA	TTTCCCTTTG	GGACAAAAG3	GA3A3AGGCA	ATAACTGTGC	TGCCAGAGTT	2700
AAATTTGTAC	GTG3ASTAGC	AGGAATCAT	TT3CTGAAAA	T3AAAAACAGA	GATGATGTTG	2760
TAGAG3CTCT	GAAGAGAGCA	AAGAAATTTT	GAAATTTG3G	CTATCAGCTA	TGGAAGAGAG	2820
TGCTGAAC3G	GAAAACAAAA	GAAGTATTGA	CAATTG3TAT	GCTTGTAATG	GCACCGATTT	2880
GAACGCTTGT	GCCATTGTTT	ACCAGCAGCA	CTCAGCAGCC	AAGTTTGGAG	TTTTGTAGCA	2940
GAAAGACAAA	TAA3TTAGGG	ATTTAATATC	CT33CCAAAT	G3TAGACAAA	ATGAACTCTG	3000
AGATCCAGCT	GCACAGGGAA	G3AAG33AAG	AC333AAGAG	GTTAGATAG3	AAATACAAGA	3060
GTCAG3AGAC	T33AAGATGT	TGTGATATTT	AA3AACACAT	AGAGTTG3AG	TAAAAGTGTA	3120
AGAAAAC3TAG	AAG33TAAGA	GACCG3TCAG	AAAG3TAG3CT	ATTTGAAGTT	AACACTTCAG	3180
AG3CAGAGTA	GTTCT3AAT3	GTAACAAGAA	ATT3AGTGT3	CCTTTGAGAG	TAGGTTAAAA	3240
AA3AATAG3C	AAC3TTTATTG	TAGCTACTTC	TG3AACAGAA	GATTGTCATT	AATAGTTTTA	3300
GAAAAC3TAAA	ATATATAGCA	TACTTATTTG	TCAATTAACA	AAGAACTAT	GTATTTTTTA	3360
ATGAGATTTA	ATGTTTATT3	TAG				3383

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11464 base pairs
- (E) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION: 1..3
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 4..82
- (C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron
 (B) LOCATION: 83..1453
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 1454..1465
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 1466..4848
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 4849..4865
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 4866..4983
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 4984..6317
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 6318..6451
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 6452..11224
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 11225..11443
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: 3'UTR
 (B) LOCATION: 11444..11464
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA	48
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala	
-35 -30 -25	
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGG CTAATGCCAT	98
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	
AGAACAAATA CCAGSTTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAACT	158
ATTAAGTGAC TCTTTSTGTC ACCAAATTTT ACTGTAATAT TAATGGCTCT TAAAAAATA	218
GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCAAGT	278
GTCCTGGCAC TTTAATCAGT AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA	338
AAATCCCAST TTTGATGGGA AAATCCCAST TTTCAATTGGA TTTCCATGCG AAAAATCCCA	398
GTACAAAACCT GGGTSCATTC ABGAAATACA ATTTCCCAAA GCAAATTGCG AAATTATGTA	458
AGAGATTCTC TAAATTTAGA GTTCCSTGAA TTACACCATT TTATGTAAAT ATGTTTGACA	518
AGTAAAAATT GATTCTTTTT TTTTTTTTTT GTTGCCCAGG CTGGAGTGCA GTGGCACAAT	578
CTCTGCTCAC TGCAACCTCC AACTCTCTGG TCAAGCAAT TCTCCTGCTT CAGCCTTCTG	638
AGTAGCTGGG ACTACAGGTG CATCCCGCCA TGCCCTGGTA ATTTTGGGGT ATTTTACTA	698
GAGACAGGCT TTTGGCATGT TGTCCAGGCT GGTCTTGGAC TCCTGATCTC AGATGATCCT	758
CCTGCTCGGG GTTCCCAAAG TCTCTGGATT ACAGGCATGA ACCACCACAC ATGGCCATAA	818
AATTGATTCT TATGATTAAT CTCTCTGAA CAATTTGCTT TCATTTGAAA GTTTGCTTCT	878
ATTTGAAACC TTCATTTAAA AGCTTGAGCA ACAAAGTGAG ACCCCATCTC TACAAAAAAC	938
TGCAAAATAT CTTCTTGACA CTTCTACCT TCTGTGGAGG CTGAAACAGG AGGATCACTT	998
GAGCCTAGGA ATTTGAGCCT GAGTGAGCT ATGATCCAC CCCTACACTC CAGCCTGCAT	1058
GACAGTAGAC CTTGACACAC ACACACAAA AAAAACCTTC ATAAAAAATT ATTAGTTGAC	1118
TTTTCTTAGG TGACTTTCCG TTTAAGCAAT AAATTTAAAA GTAAAACTC TAATTTTAGA	1178
AAATTTATTT TTAGTTACAT ATTGAAATTT TTAAACCCTA GGTTTAAGTT TTATGTCTAA	1238
ATTACCTGAG AACACACTAA GTCTGATAAG CTTCAATTTA TGGGCCTTTT GGATGATTAT	1298
ATAATATTCT GATGAAAGCC AAGACAGACC CTTAAACCAT AAAAATAGGA GTTCGAGAAA	1358
GAGGAGTAGC AAAAGTAAAA GCTAGAATGA GATTGAATTC TGAGTCGAAA TACAAAATTT	1418
TACATATTCT GTTTCTCTCT TTTTCCCCCT CTTAG CT GAA GAT GAT G GTAAA	1470
Ala Glu Asp Asp Glu	
-10	
GTAGAAATGA ATTTATTTTT CTTTGCAAA TAAGTATCTG CTTGAGACAC ATCTATCTCA	1530
CCATTGTCAG CTGAGGAAAA AAAAAAATG TTCTCATGCT ACCAATCTGC CTTCAAAGAA	1590
ATGTGGACTC AGTAGCACAG CTTTGGAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC	1650

CTCTA3CAAA	AGATGCTTCT	CTATGCTTCA	AAAAATTCTC	CAGCTCTTAT	AATCTACAAA	1710
ATAGACTTTT	CCTGTTTCAT	TGGTCTTAAG	ATTAGCATGA	AGCCATGGAT	TCTGTTGTAG	1770
GGGGAGCGTT	GCATAGGAAA	AAGGGATTGA	AGCATTAGAA	TTGTCCAAAA	TCAATAACAC	1830
CTCCTCTCA3	AAATGCTTTG	G3AAGAAGCC	TGGAAGGTTT	CGGGTTGGTG	GTGGGGTGGG	1890
GCAGAAAATT	CTG3AAGTAG	AG3AATAGG	AATGGGTGGG	GCAAGAAGAC	CACATTACAG	1950
G3CCAAAAGC	TGAAAAGAAC	CATGGCATTT	AT3ATGAATT	CAG3GTAATT	CAGAATGGAA	2010
GTA3A3TAGG	A3TA33AGAC	TG3TGAGAGG	AG3TAGAGT3	ATAAACAGG3	TGTAGAGTAA	2070
GAG3TTCTCT	CACCCCAAGA	T3TGAAATTT	GGACTTTTAT	TT33AGATAA	TAGGGTTAAT	2130
TAAGCACAAAT	ATGTATTAGT	TAG33TAAAG	ATTAGTTTGT	TGTAACAAA3	ACATCCAAA3	2190
ATACAGTAGC	TGAATAAGAT	AGAGAAATTT	TCTCTCAAA3	AAA3TCTAA3	TAGGCAGCTC	2250
AGAAGTAGTA	TGGCTGGAAG	CAACCTGATG	ATATTG33AC	CCCAACCTTT	CTTCAGTCTT	2310
GTACCCATCA	TCCCCTAGTT	GTTGATCTCA	CTCACATAGT	TGAAAATCAT	CATACTTCCT	2370
G33TTTCATAT	CCCATTTATC	AAG3AAGGGT	CAAGAGAAGT	CA33CTCATT	CCTTTCAAA3	2430
ACTCTAATTG	GAA3TTAAAC	ACATCAATCC	CCCTCATATT	CCATTGACTA	GAATTTAATC	2490
ACATG3CCAC	ACCAAGTGCA	AG3AAATCT3	GAATAATATA	TCTTTATTCC	AG3TAGCCAT	2550
ATGACTCTTT	AAAATTCAGA	AATAATATAT	TTTTAAATA	TCATTCTG3C	TTTGGTATAA	2610
AGAATT3ATG	GTGTGGGGTG	AGGAGGCCAA	AATTAAGGGT	TGA3AGCCTA	TTATTTTATG	2670
TATTACAAGA	AATGATGGT3	TCATGAATTA	AG3TAGACAT	AGGGGAGTG3	TGATGAG3AG	2730
CTGTGAATG3	ATTTTAGAAA	CACCTGAGAG	AATCAATA33	ACATGATTTA	GG3TTGGATT	2790
TGGAAA33AG	AAGAAAGTAG	AAA3ATGAT	GCCTACATTT	TTCACCTAG3	CAATTTGTAC	2850
CATTCA3T3A	AATAG33AAC	ACAGGAGGAA	GACGAGGTTT	TGGTGTATAC	AAAGAGGAGG	2910
ATGGAT3ACG	CATTTCGTTT	TGGATCTGAG	ATGTCTGT3G	AACGTCCTAG	TGGAGAT3TC	2970
CACAAACTCT	TCTACATGTG	GTTCTGAGTT	CAGGACACAG	ATTTGGGCTG	GAGATAGAGA	3030
TATTGTAG3C	TTATACATAG	AAATGGCATT	TGAATCTATA	GAGATAAAAA	GACACATCAG	3090
AGAAATGT3	TAAAGTGAGA	GAG3AAAAGC	CAAGTACTGT	GCTGGGGG3A	ATACCTACAT	3150
T3AAAG3ATG	CAGTAGAAAG	LAGCTAATAA	ACAACAGAG3	GCAGACTAAC	CAAAAGGGGA	3210
GAAGAAAAAC	CAAGAGAATT	CCAG3GACTC	CCAG3AGAG3C	ATTTCAAGAT	TGAGGGGATA	3270
G3TGTT3T3T	TGAATTTTGC	AGCCTTGAGA	ATCAAG3GCC	AGAACACAGC	TTTTAGATTT	3330
AGCAACAAG3	AGTTTGGTGA	TCTCAGTGAA	AGCAGCTTGA	TGGTGAAAT3	GAGGCAGAGG	3390
CAGATT3C3A	TGAGT3AAAC	AGT3AATGG3	AGGTGAAGAA	ATGATACAG3	TAATTTCTGC	3450
TAAAG3CTTG	GCTGTTAAAA	GGAG3AGAGA	AACAAGACTA	GCTGCAAA3T	GAGATTGGGT	3510
TGATGGAG3A	GTTTTAAATC	TCAAAATAAA	GAGCTTTGTG	CTTTTTTTGAT	TATGAAAATA	3570
ATGTGTTAAT	TGTAACATAA	TGAG3CAAT3	AAAAA3ATA	ATAATATGAA	AGATAAAAAAT	3630
ATAAAAAACA	CCCAGAAATA	ATGATAGCTA	CCATTTTGTAT	ACAATATTTT	TACACTCCTT	3690
TCTAT3TATA	TATACAGACA	CAGAAATGCT	TATATTTTTA	TTAAAAGG3A	TTGTACTATA	3750
CCTAAGCT3C	TTTTTCTAGT	TAGT3ATATA	TATGGACATC	TCTCCAT33C	AACGAGTAAT	3810
T3CAGTTATA	TTAAGTTTAT	GATATTTTAC	AATAAG3GCA	TATCTTTGCC	CTTTTTATTT	3870
AATCAATTCT	TAATTT3GT3A	ATGTTTGTTT	CCA3TTTGTG	GTTGTTATTA	ACAATGTTCC	3930
CATAAGCATT	CCTGTACACC	AATGTTTACA	CATTTTGTCTG	ATTTTTTTCTT	CAGGATA3AA	3990
CCAG3AG3T	AGAATTGCT3	G3TT3ATAGA	AGAGAAAGGA	TGATTGCTAA	ATTAAGCTTT	4050
CAGTAGAG3G	TACAT3CC3A	GCACAAAT3G	GATCAGCCCT	AGATAACCAG	AATGSCACTT	4110
TCTCATTTCC	CCTT33GACA	AAAG3GAGAG	AG3CAATAAC	TGTGCTG3CA	GAGTTAAATTT	4170
T3TAC3T33A	GTAG3AG3AA	ATCATTTTGT	GAATAATGAA	ACAGAGATGA	TGTTGTA3AG	4230
GTCT3A33A	GAGCAAA3AA	AATTTGAAAT	TGCGGCTATC	AGCTATG3AA	GAGAGTGCTG	4290
AATG33AAAA	CAAAAGAA3T	ATT3ACAATT	G3TATGCTTG	TAATGG3ACC	GATTTGAA3G	4350
CTTGT3CCAT	TGTTCA3CAG	CAGCACTCAG	CAGCCAA3TT	TGGAGTTTGT	TAGCAGAAAG	4410
ACAAATAA3T	TAGGGATTTA	ATATCTG33C	CAAATG3TAG	ACAAAATGAA	CTCTGAGATC	4470
CAGCTGCACA	GGGAAG3AA3	GGAA3AC33G	AA3AG3TTAG	ATAG3AAATA	CAAGAGTCAG	4530
GAGACTG3AA	GATGTT3T3A	TATTTAAGAA	CACATAGAGT	TGGAGTAAAA	GTGTAAG3AA	4590
ACTAGAG33G	TAAGAGAC3G	GTCA3AAAT3	AG3CTATTTG	AGTTAACA3C	TTCAGAG3CA	4650
GAGTAGTTCT	GAAT33TAA3	AAGAAATTT3A	GT3T3CCTTT	CAGAGTAG3T	TAAAAAATAA	4710
TAG3CAACTT	TATTGTAG3T	ACTTCTG3AA	CAGAA3ATTTG	TCATTAATA3	TTTTAGAAAA	4770
CTAAATATA	TAGCATACTT	ATTGTCATAT	TAACAAAGAA	ACTATGTATT	TTTAAT3AG	4830
ATTTAAT3T	TATT3TAG	AA AAC CTG	GAA TCA GAT	TAC TTT GGC AAG CTT		4880
		Glu Asn Leu	Glu Ser Asp	Tyr Phe Gly Lys Leu		
		-5		1	5	
GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT				GAC CAA GTT CTC TTC		4928
Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn				Asp Gln Val Leu Phe		
		10	15	20		
ATT GAC CAA GGA AAT CGG CCF CTA TTT GAA GAT				ATG ACT GAT TCT GAC		4976
Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp				Met Thr Asp Ser Asp		
		25	30	35		
TGT AGA G GTATTTTTT TTAATTTGCA AACATAGAAA				TGACTAGCTA CTTCTTCCCA		5032
Cys Arg Asp						
		40				
TTCTGTTTTA CTGCTTACAT TGTTCCGTGC TAGTCCCAAT				CCTCAGATGA AAAGTCACAG		5092
GAGTGACAAT AATTTCACTT ACAGGAAACT TTATAAG3CA				TCCAC3TTTTT TTAGTTG3GG		5152

TAAAAAATTG	GATACAATAA	GATATTGCTA	GGGGTCTATG	CTCTCTGAGG	CTGCTTTTGA	5212
ATCAACCAAT	CCTTTTATTG	GATTGCATTA	ACTGTTTTAA	ACCTCTATAG	TTGGATGCTT	5272
AATCCCTGCT	TGTTACAGCT	GAAAATGCTG	ATAGTTTTAC	AGGTGTGCTG	GCATCTATCT	5332
GTAATCCTAG	CTACTTGCGA	GGCTCAAGCA	GGAGGATTG	TTGAGGCTAG	GACTTTGAGG	5392
CTGTAGTACA	CTGTGATGCT	ACCTGTGAA	AGGCACTGCA	CTGCAAGCTG	GGTGATATAC	5452
AGACCTTGT	TCTAAAAATA	AAAAAAAATA	AAAAAAAATA	CTTAAAGAAA	GAAATTGATC	5512
AAGTCTACTG	TGCTTTCCAA	AACATGAAAT	CTAAATATCA	AAGTTAAGCT	GAGTTGAAAG	5572
AGTGAATGTG	CATTCTTTAA	AAATACTGAA	TACTTAACCT	AACATATATT	TTAAATATTT	5632
TATTTAGCAT	TTAAAGGTTA	AAAACAATCT	TTTAGAATTC	ATATCTTTAA	AATACTCAAA	5692
AAAGTTGCGG	CGTGTGTGTT	GTAATACACA	TTAAACTGTG	GGGTTGTTTG	TTTGTGTTG	5752
ATGCAGTTTC	ACTCTGTGAC	CCAGGCTGAA	TTGCACTGCA	GTGCACTGCT	GTGATCTCGG	5812
CTCACTACAA	CCTCCACCTC	CCAGGTTCAA	GGGATTCTCA	TGCTTCAATC	TCCCGAGTAG	5872
GTGGGATTA	AGGCATGAC	CACTTACAC	CGGCTAATTT	TTTATTTTTT	AGTAGAGCTG	5932
GGGTTTCAC	ATGTTGGCCA	GGCTGCTCT	AAACCCCTAA	CCTCAAGTGA	TCTGCTGCTC	5992
TCAAGCTCCC	AAACAAACAA	ACAACCCAC	AGTTTAAAT	GTGTTACAAC	ACACATGCTG	6052
CAACTTTTAT	GAGTATTTTA	ATGATATAGA	TTATAAAGG	TTGTTTTTAA	CTTTTAAATG	6112
CTGGGATTA	AGGCATGAGC	CACTGTGCTA	GGCTGAACT	GTGTTTTTAA	AAATGCTCTG	6172
CCAGCTGTAC	ATAGTCTCCT	GGAGCTGCT	CAAGTCTCAA	AGTGGGAAAC	GGTGTATTAA	6232
GACTATCTCT	TTGGTTAAAT	TTGCTAAAT	GTTCCTGTCT	AAAGATTCTT	CTAAGTAGAG	6292
TTCTCATTTA	TTATATTTAT	TTTCA	AT AAT GCA CCC	CGG ACC ATA TTT ATT		6343
		Asp	Asn Ala Pro	Arg Thr Ile Phe Ile		
		40		45		
ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT	ATG GCT GTA ACT ATC	6391				
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly	Met Ala Val Thr Ile					
50	55	60				
TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC	TGT GAG AAC AAA ATT	6439				
Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser	Cys Glu Asn Lys Ile					
65	70	75				
ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTTT	CAATCATGTT AATATATCA	6495				
Ile Ser Phe Lys						
ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA	AGTAATGTAA TTAGAAACT	6556				
CAAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA	ACAAGAAGCA GAGAACCATT	6616				
AAAGTGAATA CTTACTAAAA ATTATCAAA	TCTTTACCTA TTGTGATAAT	6676				
CTGAGCCTCT CACAGGGGAA GAGGAGATAC	AACACTTGTT TTATGACCTG	6736				
ACAATCAGTC TTTATACAAA TAATAATGTA	GAATACATAT GTGAGTTATA	6796				
TAACTGTGA CTTTCCAGAA TGATTTCTGT	TATGAGAAAT GAAGCTAATT	6856				
ATTTCTACAC CTTTGTAAAT TATGATAATA	TTTTAATCCC TAGTTGTTTT	6916				
CTTAGCTTAA GTCTTAGACA CAAGCTTCAG	CTTCCAGTTG ATGTATGTTA	6976				
TAATCTAATT GAATAAAGT TATGAGATCA	GCTGTAAAG TAATGCTATA	7036				
AGCCAGGTAT AAGTATTTT	TGGCTCTCTA TTTTCTCTA	7096				
CTATTATTTT TCTCTATTTT	CTCCATTATT GTTAGATAAA	7156				
AGACTGAGGC AGTAAGAGTA GGCAGGGAAT	CTTACAAAT GGCAGTGT	7216				
TTCCATGTCA TGAAGACTCT TTTTGAATGG	AGATTTGCTA ATAAATATCC	7276				
CCACCCACTC CCACTGAAA GACAGTTAGG	ATATGACCTT AGTGAAGSTA	7336				
ACTTGGTAGG GAGAAAAAG CCACCTCTAA	ATATAATCA AGTAAAGACA	7396				
AACAGATACA GCGCCCTGAC AAATCCCTCA	GCTATCTTCC TCCAAACAGA	7456				
TTGAGGTGAC AATTTGAGT CCCCATTCTA	GACCTGACAG GCAGCTTAGT	7516				
GCATAAGAGG CCTGGGATGG AAAGGTAGGG	TGGAAAGGCT TAAGCATGCT	7576				
AACATAATTA GAAGGGAGG AGATGGGCAA	GCTCAGCTA TGTGGGATAG	7636				
AGGTGAGAG AGGAGACTG GTGAAATGT	TAAAGAGATG GAAATAATGC	7696				
GGCAAAATCCA TATTTGGGGG AGCTGAAAT	TTATTTCAAT TTGATGGCTT	7756				
AAAGAATGTG GCTGGGCTTG GTGGGCTGCA	CCTGTAATCC CAGCACTTTG	7816				
GGGGGGGCT CACCTGAAGT CAAGAGTTCA	AGAAGAGCTT GACCAACATG	7876				
ATCTCTACTA AAAATACAAA ATTAGCTGGG	CGTGGTGGCA TATGCTGTGA	7936				
CTGGGGAGGC TGAGGCAAG GAATCTTTTG	AACCGGGAG GCAGAGGTTG	7996				
AGATGCTGCC ATTGCTCTCC AGTCTGGCA	ACAAGAGCAA AACTGCTCTT	8056				
AAAAAAAAG TGAAATTAA CAAAGGCTAT	AGCTTAATAA TTTAATAGTG	8116				
GGGGGGGGGG TGGCTGGAAG AGATCTGTGT	AAATGAGGGA ATCTGACATT	8176				
CAGCATCATA GCAAACTCTC TTCTGGAAG	AATCAATAA ATATTAGTTG	8236				
GAGAGTGAGG GCTGGAGTAG GACCACTTTT	AGCTCTGTG TTTAATCTCT	8296				
CTAATAAGCA TCTTAGCAAT GGTATAAAA	GTGGCTAGG TTCTAGATAA	8356				
CAGGGCAGGC ACAATGCTCT ATGCTATATA	TCCCAAGCACT TTGGAGAGGC	8416				
TCTCACTTGA GATCAGGAG TCAAGACAG	CCTGGGCTAG ATGGGATAC	8476				
CTAAAAAAA TACAAAAAT AGCCAGGCAT	GCTGGCATGC AACTGTAATC	8536				
GTGAGCCTGA GGCAGAAAG TGGCTTGAAG	CCAGGAGGTC TAGGCTGAG	8596				
TCGCACTACT GCACTCCAG CTGGGGGACA	GAATGAGACT TTGTCTCAAA	8656				

GATACAACAG	GCTACCCCTTA	TGTGCTCAAC	TTTCACTGTT	GATTACTAGC	TATAAAGTCC	8776
TATAAAGTTC	TTTGGTCAAG	AACCTTGACA	ACACTAAGAG	GGATTTGCTT	TGAGAGGTTA	8835
CTGTCAAGT	CTGTTTCATA	TATATACATA	TACATGTATA	TATGTATCTA	TATCCAGGCT	8895
TGGCCAGGT	TCCCTCAGAC	TTTCCASTGC	ACTTGGGAGA	TGTTAGGTCA	ATATCAACTT	8955
TCCCTGAT	CAGATTC AAC	CCCTTCTGAT	GTA A A A A A A A	A A A A A A A A A A	G A A A G A A A T C	9015
CCTTTCCCT	TGGAGCACTC	AAGTTTCACC	AGTTGGGBCCT	TTCCAAATTC	GGGGTTCTCC	9075
AAGGTCATTG	GGATTGCTTT	CACATCCATT	TGCTATGTAC	CTTCCCTATG	ATGGCTGGGA	9135
GTGCTCAACA	TCAAAACTAG	GAAAGCTACT	GCCCAAGGAT	GTCCCTACCT	CTATTCTGAA	9195
ATGTGCAATA	AATGTGATTA	AAGAGATTGC	CTGTTCTACC	TATCCACACT	CTCGCTTTCA	9255
ACTGTAACCT	TCTTTTTTTC	TTTTTTCTTT	TTTTTCTTTT	TTTTTGAAAC	GGAGTCTGCG	9315
TCTGTGCCC	AGGCTAGAGT	GCAGTGGCAC	GATCTCAGCT	CACTGCAAAC	TCTGCTCCC	9375
GGGTTCAAGC	CATTCTCTCG	CCTCACCTTC	CCAAGCACT	GGGACTACAG	GCGCTGCCA	9435
CCATGCCCAAG	CTAATTTTTT	GTATTTT TAG	TAGAGACGGG	GTTTCACCCT	GTTAGCCAGG	9495
ATGCTCTGCA	TCTCCTGAA	TTGTGATCCG	CCCGCCTCAG	CCTCCCAAAG	TGCTGGGATT	9555
ACAGGCTGGA	GCCATCGCAC	CCGCTCAAC	TGTAACCTTC	TATACTGGTT	CATCTTCCCC	9615
TGTAATGTTA	CTAGAGCTTT	TGAAGTTTTG	GCTATGGATT	ATTTCTCATT	TATACATTAG	9675
ATTTCAAT	AATTCACAA	TGATGCCAC	AGCTTAGGCT	CTCTTCTTAA	ATTGTATATT	9735
GTAGACAGCT	GAGAAAGTGG	GTGCCAATAG	GGAAGTATG	TTATACTTTC	ATCAACTTAG	9795
GACCCACACT	TGTTGATAAA	GACCAAGGT	CAAGAGTTAT	GACTACTGAT	TCCACAAGT	9855
ATTGAGAAAT	TGGAGATAAT	CCCGTGACCT	CTGCTATCCA	GAGTCTTTCA	GGCATCTTTG	9915
AAGGATGAA	AAATGCTATT	TAAATTTTGG	AGTTTCTCT	ATCAGTGCTT	AGGATCATGG	9975
GAATCTGTG	TGCTCATGAG	CCAAATTTAA	GTCCAAACAA	TCTACTGGTT	CCAGGATTAA	10035
CATGGAAGAA	CCTTAGGTGG	TGCCACATG	TTCTGATCCA	TCTGCAAAA	TAGACATGCT	10095
GCACTAACAG	GAAAGTGCA	GACAGCACTA	CCAGTTGGAT	AACCTGCAAG	ATTATAGTTT	10155
CAAGTAATCT	AATCAATTTCT	CACAAGGCC	TATTTCTGTA	CTGAAACATA	CAAGAACTG	10215
CATTGCTCT	TCTAAGGCG	GACCCAGCCA	AGGAGACCAT	ATTCAAGACA	GAAATTCAG	10275
ACTACTATG	AATGAGTG	CTTGACAGG	AGACAGAGT	CAAGGACTGC	CAACTGAGCC	10335
AATACAGCA	GCTTACAG	GAACCCAGG	CCTAGCCCTA	CAACAATTAT	TGGGTCTATT	10395
CACTGTAACT	TTTAATTTCA	GGCTCCACTG	AAAGAGTAAG	CTAAGATTCC	TGGCACTTTT	10455
TGTCTCTCT	ACAGTTGCT	CAGAAATGAG	AACTGGTCAG	GCCAGGCATG	GTGGCTTACA	10515
CCTGGAATCC	CAGGCTTTG	GGAGGCCGAA	GTGGAGGCT	CACTTGAGGC	CAGGAGTTCA	10575
GGACCACTT	AGGCAACAAA	GTGAGATACC	CCCTGACCCC	TTCTCTACAA	AAATAAATTT	10635
TAAAAATTAG	CCAAATGTG	TGGTGTATAC	TTACAGTCCC	AGCTACTCAG	GAGGCTGAGG	10695
CAGGCGGATT	GCTTGAGCCC	AGGAATTCAA	GCTGCAAGT	AGCTATGATT	TCACCACTGC	10755
ACTTCTGCT	GGGCAACAGA	GCGAGACCTC	GTCTCAAAGT	AAAGAGAAA	AGAACTAGA	10815
ACTAGCTTAA	GTTTGTGGGA	GGAGGTCATC	ATGCTCTTTA	GCGGTGAATG	GTTATTATAG	10875
AGGACAGAAA	TTGACATTAG	CCCAAAAAGC	TTGTGGTCTT	TGCTGGAATC	CTACTTAATC	10935
TTGAGCTAAT	GTGACACCA	CTCAATGGGA	GAGGAGAGAA	GTAAGCTGTT	TGATGTATAG	10995
GGGAAACTA	GAGGCTGGA	ACTGAATATG	CATCCCATGA	CAGGGAGAAT	AGGAGATTCC	11055
GAGTTAAGAA	GGAGAGGAGG	TCAGTACTGC	TGTTCCAGAGA	TTTTTTTTTAT	GTAACCTCTG	11115
AGAAGCAAAA	CTACTTTTGT	TCTGTTTGT	AATATACTTC	AAACCAAACT	TCATATATTC	11175
AAATTGTTCA	TGTCTGAAA	TAATTAGGTA	ATGTTTTTTT	CTCTATAG	GAA ATG AAT	11233

85
Glu Met Asn

CCT	CCT	GAT	AAC	ATC	AAG	GAT	ACA	AAA	AGT	GAC	ATC	ATA	TTC	TTT	CAG	11281
Pro	Pro	Asp	Asn	Ile	Lys	Asp	Thr	Lys	Ser	Asp	Ile	Ile	Phe	Phe	Glu	
90 95 100																
AGA	AGT	GTC	CCA	GGA	CAT	GAT	ANT	AAG	ATG	CAA	TTT	GAA	TCT	TCA	TCA	11329
Arg	Ser	Val	Pro	Gly	His	Asp	Asn	Lys	Met	Gln	Phe	Glu	Ser	Ser	Ser	
105 110 115																
TAC	GAA	GGA	TAC	TTT	CTA	GCT	TGT	GAA	AAA	GAG	AGA	GAC	CTT	TTT	AAA	11377
Tyr	Glu	Gly	Tyr	Phe	Leu	Ala	Cys	Glu	Lys	Glu	Arg	Asp	Leu	Phe	Lys	
120 125 130 135																
CTC	ATT	TTG	AAA	AAA	GAG	GAT	GAA	TTG	GGG	GAT	AGA	TCT	ATA	ATG	TTC	11425
Leu	Ile	Leu	Lys	Lys	Glu	Asp	Glu	Leu	Gly	Asp	Arg	Ser	Ile	Met	Phe	
140 145 150																
ACT	GTT	CAA	AAC	GAA	GAT	TAG	TATTAA	AATTT	CATGC	C						11464
Thr	Val	Gln	Asn	Glu	Asp											
155																

(2) INFORMATION FOR SEQ ID NO:14:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28994 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: 5'UTR

(B) LOCATION: 1..15606

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 15607..15685

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 15686..17056

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 17057..17068

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 17069..20451

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 20452..20468

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: mat peptide

(B) LOCATION: 20469..20586

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 20587..21920

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: mat peptide

(B) LOCATION: 21921..22054

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 22055..26827

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: mat peptide

(B) LOCATION: 26828..27046

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: 3'UTR

(B) LOCATION: 27047..28994

(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACTTGCCCTTA	AAAGCTTTTGC	ATAGGTAGAC	AACATTAGAT	TAATTTTCCTT	GCTCACATCT	60
GTTCAAGAAA	AATCATTTTAA	GTTATAAAAT	ATAACAAACC	TTCTGCATTA	TAAGACTGAT	120
GTTTAGAAAT	ATAAACATTT	TATACATCAC	CATTTAAATC	TTTCTCCAAG	GCTTCATCTT	180
TATAAAATAG	TCCGAAAATT	TCAGABAAAG	ATGAATCTGA	TTTTCCAAGA	GAGGACAGCT	240
GTGGACTATC	TGGTACTGSA	GACTAAATAA	AGAAAGCAGG	TACAGTCAAT	AAGATCTTCA	300
GGACATATAC	ATTTTGTTTTA	TTAAGAAAAA	GCAAATAAAA	CATTTTTTCAG	AAAAAGGCAA	360
ACATGCTAGA	AAGCATATGA	CTTASTCATT	TGAGTTTTTTA	TTATTAAGGA	AATTTACAGG	420
CCCAAGAAAC	ACCTTGCTCA	ATATATTAAA	TTTTATTTTG	GTTTTCAACT	AGACTTTGCT	480
TTTCATTTGT	TTGTTTTTBT	GAAABTTCT	CGCTCTGTCA	CCTAGGCCAA	AGTGTAGTGA	540
CACAATCTTA	GCTCACTGTA	GCTCTCTAGA	TTCAAGTGAT	CCTCCTGTCT	CAGACTCCTG	600
AGTAGCTAGG	ACTACAGGAA	CATTCCACCA	TGCCCAGCTA	ATTTTGTTTT	GTTTTGTTTT	660
GTTTTTCAGG	ACAATGTATT	GCAGGTTTSC	CCAGGCTGAT	CTGAAACTCT	TAGCTCTCAA	720
CGATACTCCT	GCTCAGCCT	CCCAAAGCAC	TAGGATTACA	GACATGAGCC	AATGCGCCCA	780
GCCTTAAATT	AGACTTTAAA	TGTGTTTTTA	AACTCCTGTT	GAAAAAGCGT	CTGSTATCTT	840
GAACCACTAG	ATGTTTTTCAT	ABCAATGAA	CTAAACTGTA	ATTTAGACAG	TAGCCAAATG	900
CTTGTAATAT	TTTGCTAAAT	AATATAATCT	TCAAGGGAGC	AAATCATGTC	CCAAATGCAA	960
AAGATCAACT	GCTGCGGCA	GTAATAAAAG	ACAGGATACT	GTGCTCTTTA	AAAGGTCAGT	1020
AACTATAGTA	CCTAGTTATC	TTACTTATCA	CAGCAAAATA	ATTACATAAA	ATCCTATGGA	1080
TCATAAAGGC	ACAGACTCAC	TTCTGTCTCT	AGATCTCAAG	CTACCAAAAA	GAAATCTCCC	1140
AATAGTTTCT	TGGAGGCTCA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCTCTC	1200
TATGTTGTAG	GACAATCCTA	GCTCTGGCA	TACGAATACA	TTAAATCCCA	CTTATCTATA	1260

GAGCTTTTCTT	AAAGGGAAGA	AATTTGAGTA	GTATGTAAAA	CAGAATAAAA	GATTAAGGCT	1320
CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTTCTA	TTATCAGGCT	1380
TTACAAGGTG	ATTTGCCATC	ATATTCGSA	GGCACCAGCT	ACAAAGCTTA	GAACAATGCC	1440
AGATTTAGGT	ACAAAATCCA	TGCTACAAGC	TCTCTGGAAT	CCTTCCCTGT	TTCCCACTCC	1500
TACTGCTGAT	GTTAATTTAG	ACTGTCTATTA	TCTGTCACTT	TCCTAAACTC	AATTTCTCCC	1560
TCCTCTAAAT	CATTCTATCA	ACTGCTATTT	GGGTAATCTT	TCAAAACCTT	GATTACTGCA	1620
TTCCCTTTAAC	TCAAAACCTT	TCATTGTCTC	AGAATAAGTT	GAAATTTCCAT	GATATGGGCT	1680
TCAGGCTCCT	GTATTATCTG	GTGCAAGGCT	ACTAGTCCCA	TCATTTTCAA	CTACTCCTCT	1740
CTATGTACTT	AGCCAAATGA	GTCTCTCTGG	CAATTCTGCC	TTGTTTCAAG	ACTGGGCTCAG	1800
TTAAGATTCT	TTTATCTTCG	GGGGGGGGGG	CTGGCTCAGG	GCTGTAATCC	CAGCACTTTG	1860
G3AAGCTGAG	GCAGGAAGAT	CACCTGAGGT	CGGAGTTTGG	AGACCAAGCT	GGCCAGCATG	1920
GTGAAACCTT	GTGTCTACTA	AAAATCCAAA	CATTAGCCAG	GCGTGGTGGG	AGGGGCTCTG	1980
AATCCAGCT	ACTTGGGAAG	CTGAGGTGAG	AGAATCGCTT	GAACCCAGGA	GAGGGAGGTT	2040
GCACTGAGGC	GAGATTGTGC	CATTGCACTC	CAGGCTGGGG	AACAGAGGGA	GACTCCACCT	2100
CAAAAAAATA	AAGGATTCTT	CTATCTTCA	AAAACTCTAA	TGTTTAAACA	GCTCTTACAG	2160
TTGATCTAAT	TCAATCTCAT	TTTTTACAA	TGAGAAACA	GGAACAGTGA	CGGTGGATCA	2220
AAGTACACCA	GTAAGACTGA	GCTAAATTA	AACCGAGATC	TCAGTGGAGT	CTGAGGTTAT	2280
TOCCACTGTC	CAACCTTACT	TTAAAGTAG	TTCAATTTT	ACTTTTACTT	TTCCATAAAT	2340
TOGGAAGGGA	TTTTCCCTAG	GAGTCCAAAT	GTTGAAACCT	GGAAGGATAT	AGTCTCTGTG	2400
TCTTTGAGAT	GAGGGGAGCC	CTGTCCATAT	TCAAGTTTAT	AATTGACTTT	GTTGTTTTTG	2460
AGAAACGAT	CTGATTGGG	TAACTTTAA	ACATCTGTTT	GATTAGTCC	ATAAATATG	2520
CATATATAGA	AGACAGAAAG	AGCAACAACA	AATTTGAAAG	ATGCTTGTTA	AGTAAATCT	2580
GTATCGTAG	TGTCCATTCC	TGCGAGTACC	TTTATAGTAT	GTAAGTTTAC	GTGCTGTAAT	2640
AGTATTAAATA	GTATCTAGAA	AATACTACAC	ATGACAGACA	GTGCTAACTT	TGCCTTGGA	2700
GTTGGAAAT	ACTTCAGAGA	AGCCAAACAG	CAGATTTTTT	TCTCTTCCCT	TCCCTTCTTA	2760
ATTTTCCCTT	TCCCTTCTAC	CCCTTCTCT	TCTCTCCCCA	AGTAACACTG	TGCACATATG	2820
TCAAACGAAA	ACTTATAATC	AAGTAACTGT	TTCTGCAAAA	ATAAGTTCGT	TTTCTGTGCA	2880
TGGCTCAAGG	CCTCAGCAGA	TCCAGGCTCT	GTGAGCGGGG	TGCTCTTCGT	CGTGTGCCAA	2940
ACACTGACCA	CTGCCCTGGC	TCTGCCATCT	TAGGCTTAAT	GACCTGGGCT	TTACTAAACA	3000
CTGTCCCTCT	TGCCCCATGG	AGCTGTCTCC	TTCTAGTCTT	CTCCCTCTTC	TCAACCGCAT	3060
CCTAGCCCTT	CAGGCTCATTT	CACCTCCATT	TTCCCTCACT	TCCCGGCGGC	CCTCCGCACT	3120
TCCTCCCTAC	GTGTTGTTCC	GCCCCACTAG	AGCCCTCAG	AGAAAGTTTC	CATCCTCGCA	3180
CCCTTCCCTG	TGTCCAGAGC	CTCACATTT	TCAAGGCGGC	CCATCCCTCC	AGCCCCACCC	3240
CAAGGCGAAT	GTACTTGGGG	GTATGCGGAC	CTTCTCTGCT	AGCGAAGCGG	AGGAGTGAA	3300
GACCTGCGG	GGGGGGTGGT	CGGACTTCCG	GGGTGAGGTT	GGGAAGCGGG	CGGCACTCCC	3360
AGTAGCCCTT	GCACAGTCA	CGTGCAGCTC	CTCCACGAC	CACTCCCGCC	AACTTCCGCA	3420
CCGTAGCCTC	CCAGAGCCAG	GGCCCCGCTA	AAAGGAGCTT	TTCCCGGTTT	TTCTCCCGCT	3480
CTTTCCCTCT	CACTTGGAAT	ACTCGTGAAG	CAAAAATCTC	TCCCTGCGCA	CCTGTGTGTG	3540
TTTGAACGAG	GAAAAATCT	GAAACTGGTC	AAAGAGAAAC	AAAGAGAGCT	TGCCAAAGCA	3600
AGGCGGCTGT	GTGTCCAGAG	AGCTTAGAAT	CTCAGCAAG	GAACACAAAA	TAGCACATTC	3660
ADGGCTCTTT	TTGAGTAAA	ATTTACTTGG	TTTCTTGGCA	GGAAGGTTT	AAAACCTGGT	3720
TTGAGATGG	TCTGTTGGCA	GGAAGGCTTT	AATCAAGTGT	TCCCTGCGGC	CACAGCAAG	3780
GTTTTTATAT	CCAGAGCCTC	AGTTACTGGC	CCCTCTTCT	CTTGGTGGCA	AGCAACGTT	3840
CAGAACTCAG	CCTTCTTAGA	AAATTCTTAC	CCGGGGTGGT	TCAATAAGTT	AAGTCTAATT	3900
GSCAACAGCT	ATCAAAAAGT	GTTGCATATC	ACACATGGCT	CACATAATTG	TAGCTTTGGC	3960
TCATCGGGTG	TTTTAATGGG	GAGGCTTTGA	CCTGCAATT	CAAGATATA	CATTCCAAAG	4020
TTAGCGCCAG	TTAGTGGATG	TGGAAGAAAG	AAAAAAGCAA	ATTACCTCAT	AACACAAAG	4080
TCATAAACAC	ACATCCATAA	GTTCCAGGTA	CAAAATCTTA	CATCTTAGAG	AACTATATTT	4140
AACATTTACA	TACATTACTA	AGGTTTTTTT	TTTCTTTTGT	CTTGATTAAA	TGTTAGTTAT	4200
CATTAACTCT	TGGAATTATT	CTGTGTGTGT	ATATTTATTT	GCTGTTTGTG	AAAGAGCGGG	4260
TTGTTTTAAA	TAAATTTCTA	GAAATATAG	GCTCAATGTG	TTTAATCTGA	GTTGCTAATA	4320
TTTGAATAA	TAGGCCACAT	AATACTAGCC	TAGATACTTA	TGCGGAAGTA	AGGAGTCTCA	4380
AACACTGTCT	CAGAACAATA	GCAATCTGTG	TTGAATTTTT	ACCTCTGTGT	GTAAAATGAA	4440
GGGAAAGGGA	ATGAAGTTTT	AGTTTCTCTT	AATTTTATCT	TTTATTTGTT	CAGACTCTTC	4500
AGCAGTATAA	AGTTTTTCTC	AAGTCAATA	TATTCACTTT	AAAGTCACTG	TGCTTTTATC	4560
TGATACCATG	TCTTCTCTAA	TTTGGGGGGG	CAGGTGAGAT	AGTTTTTATG	AAATAAAG	4620
ATTAAAAAT	CTTACATTTT	TAGTGTCTCT	CCTTGGTAAA	ATGTAGATTT	GTCCACTGTG	4680
TTTATCTCTT	CCTCCTTATT	ATCATGTTTG	CTTATTATT	TTTTAATGGT	TCATTAAACC	4740
CAAGGCTCTG	GGAATACTCT	ATGGAATTTA	TCTCAGAGCC	TTTCACTGCT	ATGATATTTA	4800
AACAGGTGCT	TGTCCATCTG	ATTCTTAAAG	TATTTCCAG	AAAAATGATT	CCACCTAATG	4860
CATAAATGCT	TTCATCAATG	TAAGGAAACA	CGATGGACAT	TTTATTTTAT	TTTATTTTAT	4920
AAATATTAAC	TTCCATTGCA	TAAGCTAAAT	GGTAGGAAT	AAGTGAAGAT	ATATTGTTAT	4980
CTAGAGCTTT	AAAATATTCA	AAGGGCTGTG	ATCATTATCT	CATTTAATCT	TTGAAAACAA	5040
CTCTATGAAG	TACAAAGGAC	ACTGAGACAT	TTGTTGCTCT	ATATCAAGGA	AAAAAGTGTT	5100
TGTCCCAAAA	CTTCAAAATG	TGTAAATTAC	ACATTCTGCA	TCITTTACAG	TGGAGAAAAT	5160
TCAGTGGCAA	TGGAATATTT	AAAATTAAG	CTTGTCTAGT	GTGCTGCTTC	TGATCACTAC	5220
TTGATCCCA	TTGCTGCTTT	CATGTTAATT	GGCCCAATTG	GACTCTACAG	TTGGAAGGTG	5280

AAAACCTTACT	ATTTTCAACTT	GASTCACSTA	TGTATTETTA	TCATATACTT	CTTAAAGSTA	5340
CTATTTTTTTT	TCTTCTGATA	GTACACACAC	CAAGCACTTC	CAGCCACCCT	GCCACAGACT	5400
TCCTTTGTAA	TCACTGTTGA	AGBACATBAT	GTTTTTATGA	CTTCCCGAAA	TGAAAACCCCT	5460
ATCTTGTTTT	TAAAAACAAAC	AAACCAACAA	AAAGTASTST	TTATGTAAGC	ATTTTSTTCC	5520
CTGACTCTAG	GAACCCCTCT	GTTTTTATAT	CAACTCTGTA	CTGGCAAAAC	ACAAAAACAA	5580
AATGCCACCT	TGCTAATTCC	CTTCTAGCA	AAGTAATACA	GTTTAGCACA	TGTTCAAGAA	5640
AAAAATGGCT	AAGAAATTTT	GTTTCCACTA	ATTATTTTCA	ABCTGTGAT	ATTTACACTC	5700
TGCTCTTCAA	ACGTTACATT	TTATAAGACT	ATTTTTTAAC	ATGTTGAACT	TAAGCCCTAA	5760
ATATATGTAT	CCTTAAATTT	TATTTCAAAAT	ATTTTAGSTC	AGTCTTTGCT	ATCATTCOCAG	5820
GAATAGAAAG	TTTTAACACT	GGAAACTGCA	AGTAAATATT	TGCCCTCTTA	CCTGAATTTT	5880
GGTAGCCCTC	TCCCCAAGST	TACTTTCTGT	TGCAGAAAT	GTAAAAATTA	TTACATAAAA	5940
TTCTAATGAT	GSTATCCGTC	TGCTTTGCT	CTGATACAGC	AGATAAAGAA	GTTTTATGAA	6000
AATGGAAGCT	TGTTCCACTG	AAAAATGAAAT	CTTAATGBCG	TGATCAACT	ATCCTTTGAC	6060
ACCATATTGA	GCTTGGGAGG	AAGGGGAAGT	CCTGAATGAG	GTTATAAAAT	AAAAGAAAAT	6120
ATTTGCAAAA	TSTTCTTTT	TTTAAATGTT	TACATTTTAG	AAATATTTTA	AGTGTGTGTA	6180
CATTGTAGAA	TAGACCCCAA	TAGGACTGAT	TATTCGCTAT	TGTAAATATA	CAAAAAGSTT	6240
TGTGCTGAAG	TGTGACCAGG	AAGTCTGAAA	ATGAAGAGAG	ACAATATGACA	AAAGAAGATG	6300
CTTCTAATG	ACTAAGGAGG	TGCTTTCTTA	AAGTCAAGAA	GAGATACTCA	GAAAGAGGTA	6360
CAAGTTTTT	AAGGCACACA	GCCCCAACTT	TTAGGGAAGA	AAAATTTTCA	TGAAAATAGT	6420
GATATTACAT	TAAAAGAAAT	ACTCGTATCC	TCTGCAACTT	TATTTGCACT	TCCATTGCCC	6480
TAGGAAGAG	CCTGTTTTGA	GGCGGGCCCA	AGGASTGCCG	ACAGCAATCT	CCTCCCTCCA	6540
CCTTCTTCT	CATTCTCTCC	CCAGCTTCT	GAGCCCTTTG	CTCCCTTGCC	GACTGCTTGG	6600
ACAGTCAGCA	AGGAATTTCT	TCCCAGTCCA	TTTTTCCCTC	CTGCTCTCCA	ACTCTGCTCG	6660
CTAAAGCGGC	TGCCACCTGC	TGCACTCTAC	ACAGTTCTCG	GAAAGAGGAA	GGAACTCTAG	6720
ACCTTTCCAG	TGCTTCTCT	TGCAACAAA	CTATTGTTG	CAGGTAAGAA	ATATCTTCT	6780
TCTTTATTTT	GAAAGTCAGC	CATGSCAATT	AGAGGTAAAT	AAGGTAGAAA	GCAATTGAGA	6840
GGAAATATAA	CCATCTAGCA	TCACTAGGAT	GAGCACTCAG	TATCAACATA	AGAAATATAA	6900
GCAAGTCAAG	AGTAGAATTT	TTTTCTTTTA	TCAATATATG	GAGATATCA	CTTTAGAGGA	6960
GAGGTTCTCA	AACTTTTTCG	TCTCATSTTC	CCTTTACATT	AAGCACATCA	CATGTTAGCA	7020
TAAATACAT	TTTTAATTAA	AAATAACTAT	GTACTTTTTC	AAACACAAA	AAAAGCTATA	7080
AGAGTGACAC	TTTTTTATTT	TTACAAGTGT	TTTTATGCT	TTAATAGAGG	CCATATAGAT	7140
CTGCTGGAAT	CTCATCTGCT	TTGCATTGAG	ACTACTGCAA	TATTGACACG	ANTGAGGCT	7200
CTGTTAACT	CTGTTGTACA	CTCATGAGAG	AATGCTTCAA	AAAGACAAAT	TACGTCTTAG	7260
AATTATTAGA	AATAGCTTTC	ACTTTAGGAA	CTCCTTGAGA	ATTGCTGCTT	TAGAGTGSTA	7320
AGATAAATAA	GCTTCTCTTT	AAACGGGATC	TCAAGACAGA	ATCAAGTTACA	TTAAAGGCAA	7380
ACAAAAAATT	TGCCCATGCT	TAGTCACTTT	GTGAATCTG	CCACACCTTT	GGACTGGGCT	7440
ACAATTGAT	AATATAGCAT	TCCCCGAGAT	AATTTTCTCT	CACAATTAGG	GAAGGGGCTG	7500
AATAAATATC	TCTGTTTGAA	GTTGAATAAC	AAAAATTAGG	AGCCCTTAAA	CTTTAGGGCT	7560
CCTGAAATTC	GTCTTTTTCG	CTATATTGAG	CTACTTTAGG	TTCTATTAAA	TCTTTTTCGA	7620
GCTCAGSTCT	ACTAGCTCAT	GCTTAGGATC	CTAGGCAAGC	CTGAGGCAAG	GAATTTGAGA	7680
CCAGCCAGGG	CAACACAGTC	TCTACAAAA	AATAAAAAAT	TACCTGGGTC	TGTTGGTGCA	7740
TGCTGTAGAA	ACTACTCAGG	ATGCTGAGGA	CTGCTTGAGC	CCAGGATAGC	GAATCTCTG	7800
GTGAGTTCAG	CCACTAAACA	GAGCGAGACT	TTCTCAAAA	AAACAAACAA	AAACAAACAA	7860
AACTTCTCTC	AAAATAAATT	TTTATCTGCA	ATSTTTTCTC	ATGCTCTGTC	AGATTAAATT	7920
TACTCTTTTA	CCTGATTTTC	AAAGCTCTCT	ATAATCTAAT	CGGACTTTAC	CTTGTGTTCA	7980
CTGCAAAATA	GTAGGACTGT	TCCACTACAA	TCCAAAAATC	ACAGTTTGGG	TGCASTGGCT	8040
CACTCTCTGA	ATCCCAACAC	TTTGGAAAGC	CAAGGCAAGT	GGATTGCTTC	AGCTCAGGAG	8100
TTCAAGATCA	GCTTGGGCAA	CATGBCAAAA	ACCTTCTCTC	TCAAAACAT	ACAAAAATTA	8160
GCCAGATGTC	GTAATATGTC	CCTGTAATCT	CAACTACTCA	AAAGGCTAAG	GAAGAGGAT	8220
CACTTGAGGC	CAGGAGGTC	AGGCTACAGT	GAGGCTAGTT	TACTGTGTCA	CTGCACTCCA	8280
GCTGCTGGA	TAGAGCAAGA	CCATGCTCTCA	AAAGAAAAAA	AAAGAAAAAA	AAAGAAAAAA	8340
ACATGGGCTC	ATTCASTTCA	CCCCCAACAC	AACATTSTTT	TGATTATCAC	ATAAATGCTG	8400
GTCCATTGCT	TTCTCTATCT	ATTCAAATCT	TTAAGCATTC	TTGAGATTTC	AACTCAATTC	8460
TCTTTTCAA	ACTAGGTCAT	TTAAACTACA	TCASTTCCAT	TTGATTTTTC	TTGCTTTGAG	8520
TCTACAGACT	CAAAAACAAA	AACTTAAAAA	CTTATTTTTC	AACTTTTCTG	CTACTCTCAC	8580
TTCTTCAACT	CTCACATACA	CACATTCTATA	ATAACATGCT	AGAAATGTTCA	AGGATAAAT	8640
GATTTATAGA	ACTGAAAAAT	TAGGTTTTGA	TCTTSTTCT	GTAAGATGTA	CTACTACTCT	8700
GATCTGAGST	AATTAATTAT	GTAGCATGCT	CGCTCATTTTC	ATCCCATAGC	TATTCAGAG	8760
GATTGGAATT	CCACAGCAAG	GATAAATATA	ATCATATTTG	CTTTTCAAGT	TCAAGGCTATT	8820
TTAACITTTA	ATCTAGTAGT	ATGTTTTTTC	TTGTTGTTCT	TGTTTGAAT	GGAGGCTGCT	8880
TGCTCACCC	AGGCTGAGT	GAGCTGAGCT	GAACTGGCT	CACTGCAACC	TCTGCTCAT	8940
GGGTTCAATC	AGTTATTCTG	CCTCAGTCTC	CCAACTAGCT	GGGACTACAA	GCCACATGCC	9000
ACCATGCTCT	GCTAATTTTT	GTATTTTTAG	TAGAAACAGG	GCTTCACCAT	GTTGBCAGG	9060
CTGCTCTCGA	ACTCCTGACC	TCAAGTGATC	CAGCGGCTCT	GGCTCTCCCA	AGTGTCTGGA	9120
TTACAGGCAAT	AAGCCACGCT	GCCCCAGCTA	ATAGTATSTT	TTTAAACTCT	TAGTGGCTTA	9180
ACAAATGCTG	TGTATAATA	AATATGCTAT	AAATATTTAC	TGTCTTAGAA	TTATGAAGAA	9240
GTGTTTACTA	GGCCGTTTTG	CACATATCAA	TGTTCTCTCT	CTTACAGCTT	TAATTAAGAT	9300

CTAGAATTG	AGGTTTGTAG	AGCTGGAACA	GACCTTAAAG	ATTGACTAGC	CAACTTCCTT	9350
GTCCAAATGA	GGGAACTGAG	ACCCTTAAAA	TTAAGTGAAT	TGCCCCAGAC	AAAACTG3AA	9420
CTCATGTGTC	CTAATTTCCA	TCATGAAATT	CTACCATTCA	CTA3CCTCTG	GCTAGTTTCT	9420
AAAGTATTTC	ATAACTAAAT	TTTTATTTCT	GTTTTAAAGA	ACAAATTGTC	ACTGCTTACT	9540
CCTG3GAG33	TCTTTCTGAG	GTGGTTTATA	ACTCTTAAAA	AAAAAAAAT	CAGTAGTCTG	9620
AGAAATTTAG	ACGAAATAGT	CAAGCTATTT	TTATCCAATG	GATCTATAAT	TTTCATAGAT	9660
TAGACTTAAA	TCAAAGAAAC	ACGATGAGA	AAGGAAGAGG	AAATTTAG33	AGAGGAG3AA	9720
TG333ATGAG	AACACACTAC	TTGTAATCAG	TCATAGATGT	ACT3AGAACT	AACAAGAGAA	9780
ATTGTAG3AA	AATAAGAAAT	AAGAATTCAA	AATCAACACA	TGAATAAAAA	AGAACTACT	9840
A3G3MAAAAT	G3AGAAGACA	TTAGAAAAAT	TATTTATTTT	TTAAAAATCT	GTTTTCA33C	9900
TTCCCTCCTG	TTCTTCCTCC	TTCTCATTG3	TTTTCA3GTG	GAG33AAAGT	TTAAGATGGA	9960
AAAAATATAT	ATATTCTACA	CATCCCTTTT	TAGCTCTGTT	TCAT33CAAC	AA33TTTATC	10020
ATA3CAAACT	TTTATTTCATA	CAATATTTAT	TGAGTTCTTA	CTGT3T33TA	AGCTCTTTCC	10080
AG3TGTG3AA	AATTCAGGGG	AAAAAAGACA	ACTCATTGTC	TTAAAACTCA	GATGAAAGCT	10140
GAACAGACCT	ATTTTAAATC	AAAGTAATCT	CAATTTAGGG	TAGTAAGAGC	TATTTAAGAA	10200
GCATGAATCA	GTGTGAAGCA	G3TAGSACTC	TGAG3AGAGA	ATAGTTAGCT	AGGAATGAAA	10260
GAG3AGAGAA	GTTTTCTTAG	AGGAATTAAT	AAAGCTGGGA	GTTAC3GGAT	GAAAGATGAG	10320
GCA333TTT3	CAG3CAAAAA	AAAAAAAAG	GCAG3G3AA3	GGGAA3TTCT	GGCCTGG3AG	10380
AGAGAAATAC	TGTGGCAACA	ATG3AG3AGA	GTCTGGAAGC	AAGAAAACCA	AGTAGAAGAG	10440
TATTAATAAT	GAAGATGCCA	G333TAATGA	G33CTTGATT	TAAAACTGTC	CTGTTG3AGA	10500
T33AGAG3AG	ATAACAAAT	CTGGAGACAT	TTCTGAGTTA	GAACCTACAG	TATTTATCAG	10560
ACAAG33AAA	GATTAGACAA	AGGAGTTAA3	AATGACTCCC	AGGTTTCA3T	TTGGGG3AGG	10620
TAAGTAA3AG	ATGTTTTGAA	AAGTAATGTA	TTGGATCTCT	TACCATTG3A	ACTATGTATG	10680
TGGAGCCAAA	TTAAAAATTT	TACATGTATA	TAACCTCTCC	CCCACCAACA	GTAACACTTT	10740
CCCTAACTCT	CTACTTTGTA	GCCAGACTTC	CTAAAA3AAT	AGTTT3TAGT	CAGTGTCTTT	10800
ACTTTT3333	TCCCATTTCT	TCTAGATAT	TTCTCCACCT	ACCATCTG3T	GCCTCCACTT	10860
TACCCAAACT	GTTCTACGGT	TGCCCCAAAC	TTCTTAATTG	CCAAATTCAA	TGAACAAGTT	10920
TAA3CTTATA	TGTAPATTAG	GAGCTCTACA	GTTTGATTTC	GAGCAG3CCC	TCCT3AAACC	10980
CTTTCTCTTT	CGATTTCTGT	GACACATCTC	AGATTTACAA	AACTCAACTA	ATTATTTTAC	11040
ACTT3AGCTG	TATTTTCGTT	CTTCTTTCTT	GAT3AATGAG	GTAAACACTC	AGCAATTTGC	11100
CCAA3CCAAA	AACTCGAAG	TCATCCTCA3	TTCTCTCTTC	TTCT3TTTGA	CCCACAACAG	11160
ATCAGCTGAG	AAATCCCGCT	GTTTAGTATC	TCTTGAATTC	ATTACCTTAA	TTTATAGCCT	11220
CATCAACTCT	TAATTGTTAA	AATTACTTCA	GTAGTTGTTG	TCTGACCTCT	GTCCAACTCT	11280
GTTCAATCAG	GTCCATTCTT	TTGTTCTTGG	TG3T33GT3T	GCTGTT3ACA	GAGTTTCTGCT	11340
TTTCT33333	AGGTT3AAGT	GCAGT3GAG3	ACTTCACTG3	AA333AG33C	TCCTGG3TTT	11400
AAGCAGTTCA	CCCTCCCGAG	TAGCTG33AC	TACAG3TATG	TG3CA33ACA	CC3AGCTAAT	11460
TTTGT3TTTT	CAGTAGAGAC	AGGGTTTTCAC	CAT3TTG3TC	AG3CT33TCT	CAAACTCCTG	11520
ACCTCA333A	ATCC3CC3AC	CTCAG3CTCC	CAAAGTGCTG	GAATTAAG3G	CATGAGCCAC	11580
TGCACAC33A	CCAGATCCAT	TGTTTAT3TT	GTTTCTAGAG	TGAGTTT3TA	AAACACAAAT	11640
TTGACCATAT	CTTCTCTCAA	TTTAA3TCA3	TATTTT3TTT	TTCA333AAA	AA3AGTTCAA	11700
ACTCTTTAGT	CTGCTTACAC	AAGGCTTTTG	TATTTCTGACT	CTTCTTTCCA	AGTTTTCATC	11760
AAAGTATACT	G3AAGTTACA	TTTTAT3T3A	ATTGAATTAG	GCAAG333AT	AA3AATTATA	11820
GTTTATAT33	G3AAAATGGA	AATAATGTTA	ACTCTTCCAA	ATA3TTTATC	TAGAATGACA	11880
TAAATTTCAA	GCTGTCAAGT	CAAT3AGTT	ATAAACTGTT	AACACTATTG	CCACATGCCA	11940
TGTCTCTCTA	TACTTGGTAG	AATTAT3T33	TTCTAGTCTA	TTATTATGTA	AATTAGACTT	12000
TAAATAT33C	AGAAGTTCTT	CAGACATACA	TTCTAGTATT	GTGCTTTT3A	AA3ATAATTT	12060
TAAATAATTT	TATATATGAT	AATGTTAT3C	AA3T3CTAAG	GGAT3TATT3	TTACT3CTGT	12120
GCAAA3333A	AAAAAAAANA	AACTCCAAAT	AAATATGTTG	AAACCAAGTT	TATAT3CAAG	12180
AAACCAATAT	TAAAGAGG3C	AAAGTA3CCAC	CATAATAG3C	TGT3T33AGA	CG3CAG33CTA	12240
CAAAACACTA	GTAATAAT3C	TGAGAA33TT	G3333333AA	AGAA333AAC	AATAT33TTT	12300
GTTGTTT3TA	G3TTTATGTA	CTCCAA3AAT	ATCTCTCT3C	AAAC3TTTAC	GTTTTTTTCA	12360
ARGAAAG3TT	AACTTTGG3T	GG3333AGT3	GCTCTTGG3T	GTAGT333AG	CTTTT333AG	12420
GCCAAAG333	GCAGATCA3C	TGAGGTCA33	AGTTTGAGAC	CAGCTTGACC	AAAAAT33AG	12480
AAACCC333C	CCCTCACTAC	TAAAA3AATA	CAAAATTAG3	CCG3333ACA	TG3CTTACCC	12540
CTGTGAT33C	AGCACTTTT3	GAG33333AG	CA33AGAGT3	ACCT3AGT3C	AG3AGTT33A	12600
GACCA333AT	GGAGAAACCC	GTCTCTATTA	AA3AATACAA	ATTAG333GG	CTGTGGT3T3	12660
CATGACT3TA	ATCC3AGCTA	CTCAG3AG3C	TAAGGTAGAG	AATCACTTGA	ACCCAG33AG	12720
TGGAGTTT3C	AGT3AG333A	GATCCT33CA	TT3CACT3CA	GCCT3333AA	CAAGAG33AA	12780
ACTCT3TAT3	CAAAAAACAA	AAGAAAAGAA	AG3TAA3CT	T3AACTATGT	GAGATCTTTA	12840
GAAAT3CATT	CTTTCTGTAA	AATGT3ACTA	CATTT3CTTT	ATTTATG3TA	AAAT3TT33A	12900
GG3CTCAAA3	AACCCATATT	TTCTG3TCT	CC333CT33C	TAG3CTT3GT	TCACATT3CT	12960
TCTTCTT33T	GGAA3CTCTT	CCTCT333CT	TGAAAAT33C	T3TTTCTCTT	TCAAG3TAGC	13020
ACAGT3CAT3	CTTTCTGT33	TAACCTT3T3	CAG33333AT	AAAC333AAG	AAT3AAT3TC	13080
TT3TAAAT3C	AGCTCTTA3G	TCATT3CATTA	CATTATTTT3	TAACTCTTTA	TAGATTCTTC	13140
TCTCCCACTA	GACTCTGAGT	CACTG3AGAG	TAG3AG33CA	CTCTCAATTCA	TGTGTGGTTT	13200
GGTCA3CTAC	TG3333CAAT	CCTGAT3CAT	AGTTAATGCT	CAAACTTTAA	CTG3TGAATC	13260
AGCTCAATA	TTGT3CTTCT	CTAAATCCAT	TCACTCATTTG	ACTAACTATG	TACTCAAAAT	13320

AGTAAACACC	AGTAATTTAA	TCCAAATTCCT	GCCCCATACTG	CTTGGGTACAT	TTCAGGTGAA	13380
TTAGTTTGTAT	AAATATGTGT	GTATTACATA	ATATTAAAGT	ATGTACAGAA	GATCATGCTA	13440
ATCATAATTC	ACAACTGATA	ACTAATCAAA	CATAAATGCT	CTCAGGTAA	CAAATGTCTG	13500
CCTTCTCAGT	TAATGCAGTC	ATTAACAAAC	ACCTTCTGAT	GCTGATAATA	GGGCCTTGTT	13560
CAGCAATGAA	GCCATAAAGG	TGAATAAAGA	ACATGCCCTC	GTGGAGCTCA	CAGCCTAGTC	13620
ATTATTGTTT	TGATTTTAA	TATTAATGTT	GGTTTGGGTT	TTGGTGAATA	ATGTTTAGAC	13680
TTATCTTAGT	GATCTTTTCA	TCCTTTGCTA	TATTATTTTT	CTCTAAGAGT	CTTCCTTATC	13740
CCCTCCTTTA	AAAAACTAGG	TGATAATTCT	AAATTGTAAA	TTTAAATATT	ATAAATAGCT	13800
TATAAAATTT	AATATTTATA	ATATTTAAAT	GTTTGATAAA	TATTTAAAT	TTATAATATT	13860
TAAATGTTTT	TTTAAATTCA	TTTGATACAT	AGTTTTTATT	TTATTTAAAT	GTGTTGGCCA	13920
G3CATG3TG3	CTGACACCTA	TAATCCGAGA	ACTTTGAGAG	GCCAGGTGAG	GCAAAACCATT	13980
TGAGCTCAG3	ASTTTGAGAC	CACCCCTGGC	AAAGTGGTGA	AACCCCTGCT	CTACCAAACA	14040
TATGAAAAC	TATCTGGGTC	TGGTGGGCA	CATCTGTGCT	CCAGATGGG	A3TCCAGGC	14100
TAAGATGGGA	GAATCGCTTG	AAACCCAGTG	AGAGGGGTG3	G3TGGATGTT	GCA3TGAGCT	14160
GAGATCGTGC	CACCTGCACTC	CAACCTGGST	GACAGAGTGA	GACTCCATCT	CAAAAAAATA	14220
AAATGTTATC	TAAATAAGAT	AAATTTAATA	ACTGTTGGCA	CTTAGATGAG	CATAAGGAAC	14280
TAAACCTAGA	TAAAACTATC	AAATAAGGCC	TGGGTACAGT	GACTCATGCC	TGTAATCTCA	14340
AGCACTTT33	GAGG3CAAAA	TTATACAAA3	TTAGTTGTAT	AACACCAACT	AACAACATATT	14400
TTGGGGTTAG	CTTAATTCAG	ATTAATTTTT	TTTAAACTGA	GTTTTAAATT	CCTGCTTACT	14460
CATCCATACA	TGCTAG3CCT	CATATTAT3C	TAGAAAAATT	TTGAGCACAG	ATTTATGAAT	14520
ACTCTCCT3C	ATACCATTTA	ATTTTAAAC	AAATTTTAAAT	GCAGTATATA	TGTGCTTTT	14580
TACCAACACA	TTAAATAATA	AGATCTACTG	TGAGGACTAA	ATTTCTGTAA	TTTCAAAGTA	14640
GTAATGAGTT	TAAACCATGT	CTCAAGATCT	CTGCAATAAC	TGTAGCACAA	CAGAAAATAG	14700
GTATTTCTAT	TAAATGACAGA	GTCACAACTA	CTACTAATAA	TACT3TG3TT	TGTTTTCTGC	14760
AACATATCAT	GGAGGAAATG	CTAAATTTCA	GAG3TTGGTG	AAAATACATG	TGTATTTT	14820
TCCCATGCA	AGTTGACAGA	TTTCTCACAC	TAGAGAACTC	TATTCATATA	CAAAATCTG	14880
GAAGCCTGCA	CACCTGATTT	GAAGAAGGGC	AGAAAGGAAA	AGCAAAATG3A	AGGATTTAAA	14940
TTTTTTTCAA	ATCCTGTATC	CCTTGATTTT	ACAGCAAGAT	TGTATTTATG	TATTACTTGT	15000
GTTAAAAATA	TAGTATAATC	GAGACTCCAG	ATCAAAAATC	ACCGCAGCTC	AGGGAGAAAG	15060
AGGGCCACCA	AATGCCAGAG	CCCTTCAGCC	TTCTCCCAAC	CT3CCTGTAT	CCTCAGATGG	15120
AAGCACTTTT	TTATCATTTG	TTACCTTTTA	GCTTTTGGAC	AATGAAGTCA	CAAACTTCA	15180
GCCTCTCAC	CATAGGAACC	CACCTGGTGT	AAGAGAAGGA	T3AA3CCAGT	CCTTCCTAAA	15240
GG3CACGATT	AGATGTGTTT	AT33CATCCT	CAG3TGAAAC	TATATTTATA	TTGACAAAT	15300
ATTTATATTT	CTCAAG3AAT	ACTAGAAATA	TGATTCAGTT	CAGTACTAG3	CCATTTATCT	15360
ACCTTTTATA	ATATTTGTTA	ATGAGAAAAT	GCTTTCTATC	TTCCAAATAT	CTGATGATTT	15420
GTAAGAGAAC	ACTTAAACAT	G33TATTCAT	AAG3TGAAAC	TTCT33CAT	TATTGAATGT	15480
CAAGATTGTT	CATCA3TATA	CTAG3TGATT	AAC3TGACCAC	T3AACTTGAA	G3TAGTATAA	15540
AGTAGTA3TA	AAAG3TACAA	TCATT3TCTC	TTAACAGAT3	GCTCTTTGCT	TTCATTAGGA	15600

ATAAAG	ATG	GCT	GAA	CCA	GTA	GAA	GAC	AAT	TGC	ATC	AAC	TTT	GTG	GCA	15651
Met	Ala	Ala	Glu	Pro	Val	Glu	Asp	Asn	Cys	Ile	Asn	Phe	Val	Ala	
-35								-30				-25			

ATG	AAA	TTT	ATT	GAC	AAT	ACG	CTT	TAC	TTT	ATA	G	GTAAGGC	TAATGCCATA	15702
Met	Lys	Phe	Ile	Asp	Asn	Thr	Leu	Tyr	Phe	Ile	Ala			
-20								-15				-10		

GAACAAATAC	CAG3TTCAGA	TAAATCTATT	CAATTAGAAA	AGAT3TTGTC	AGGTGAACATA	15762
TTAAGTGACT	CTTTGTGTCA	CCAAATTTCA	CT3TAATATT	AAT33TCTT	AAAAAATAG	15822
TGGACCTGTA	GAAATTAACC	ACAACATGTC	CAAG3TCTCA	GCACCTG3TC	ACACCACGTG	15882
TCCTGGTACT	TTAATCAGCA	GTAG3TCACT	CTCAG3TTG3	CAGTAAGTGC	ACATCATGAA	15942
AATCCCASTT	TTGATGG3AA	AATCCCASTT	TTGATTTGGAT	TTCCATG33A	AAAATCCAG	16002
TACAAAACCTG	GGTGCACTTCA	GGAAATACAA	TTTC3CAAG	CAAAATG3CA	AATTATGTAA	16062
GAGATTCTCT	AAATTTAGAG	TTCCGTTGAAT	TACACATTTT	TAT3TAAATA	TGTTTGACAA	16122
GTAAAAATTG	ATTCTTTTTT	TTTTTTTCTG	TT33CTAGGC	T33A3T33AG	T33CACAAATC	16182
TCTGCTCACT	GCAACCTCCA	CCTCCTG33G	T3AAGCAATT	CTCCT33CTC	AGCCTTTCTGA	16242
GTAGCTG33A	CTACAG3T3C	ATC33333AT	GCCT33CTAA	TTTTT333TA	TTTTTACTAG	16302
AGACAG33TT	TTGGTATGTT	GTCCAG3CTG	GTCTT3GACT	CCT3ATCTCA	GATGATCCTC	16362
CTGGCTCGGG	CTCCTAAAGT	GCT333BATT	CAG3CATGAA	CCACCACACA	TG3CCTAAAA	16422
ATTGATTCTT	ATGATTAATC	TCCTGTGAAC	AAITTT33CTT	CATTTGAAAG	TTTGCTTCA	16482
TTTGAAAAC	TCATTTAAAA	GCTGAG3CAA	CAAG3TGAGA	CCCCATCTCT	ACAAAAAAT	16542
GCAAAATATC	CTGTTGACAC	CTCCTACCTT	CTG3TGAGGC	T3AAG3CAG3A	GSATCACTTG	16602
AGCCTAG3AA	TTTGAGCCTG	CAGTGAGCTA	TGATC33A3C	CCTACACT3C	AGCCTG3CATG	16662
ACAGTAGACC	CTGACACACA	CACACAAAAA	AAAACCTTCA	TAAAAAATTA	TTAGTTGACT	16722
TTTCTTAG3T	GACTTTCCGT	TTAAGCAATA	AATTTAAAA3	TAAATCTCT	AATTTTAGAA	16782
AATTTATTTT	TAGTTACATA	TTGAATTTTT	TAAACCTAG	GTITAGTTT	TATGTCTAAA	16842
TTACCTGAGA	ACACACTAAG	TCTGATAAGC	TTCACTTTAT	GG3CCTTTTG	GATGATTATA	16902
TAATATTCTG	ATGAAAGCCA	AGACAGACCC	TTAAACCAT	AAAATAGGAG	TTCGAGAAA3	16962
AGGAGTAGCA	AAAGTAAAA3	CTAGAAATGAG	ATTGAATTCT	GAGTCGAAAT	ACAAAATTTT	17022

ACATATTCTG TTTCTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G GTAAAGT 17075
 Ala Glu Asp Asp Glu
 -10

AGAAATGAAT TTATTTTTCT TTGCAAACTA AGTATCTGCT TSAGACACAT CTATCTCACG 17135
 ATTGTCAGCT GAGGAAAAAA AAAAATGSTT CTCATGCTAC CAATCTGCTT TCAAGAAAAAT 17195
 GTGGACTCAG TAGCACAGCT TTGGAAATGAA GATGATCATA AGAGATACAA AGAAGAACCT 17255
 CTAGCAAAAG ATGCTTCTCT ATGCTCTAAA AAATTCTCCA GCTCTTASAA TCTACAAAAT 17315
 AGACTTTGCT TGTTTCATTG GTCTTAAGAT TAGCATGAAG COATGGATTG TGTGTAGGGG 17375
 GGAGCGTTGC ATAGGAAAAA GGGATTGAAG CATTAGAATT GTCCAAAAATC AGTAACACCT 17435
 CCTCTCAGAA ATGCTTTGGG AAGAAGCCTG GAAGGTTCCG GBTGGTGGT GGGGTGGGGG 17495
 AGAAAAATCT GGAAGTAGAG GAGATAGGAA TGGGTGGGGG AAGAAGACCA CATTACAGAAG 17555
 CCAAAAGCTG AAAGAAACCA TGBCATTTAT GATGAATTCA GGTAAATTCA GAATGGAAGT 17615
 AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGAT AAACAGGGTG TAGAGCAAGA 17675
 CGTCTCTCTA CCCCAGATG TGAATTTGG ACTTTATCTT GGAGATAATA GGGTTAATTA 17735
 AGCACAATAT GTATTAGCTA GGTAAAGAT TAGTTTGTGT TAACAAAGAC ATCCAAAGAT 17795
 ACAGTAGCTG AATAAGATAG AGAATTTTC TCTCAAAGAA ASTCTAAGTA GGCAGCTCAG 17855
 AAGTAGTATG GCTGGAAGCA ACCTGATGAT ATTGGGACCC CCAACCTTCT TCAGTCTTGT 17915
 ACCCATCATC CCTAGTTGT TGATCTCACT CACATAGTTT AAAATCATCA TACTTCCTGG 17975
 GTTCATATCC CAGTTATCAA GAAAGGGTCA AAGAAAGTCA GGTTCATTCC TTTCAAAAGC 18035
 TCTAATTGCA AGTTAAACAC ATCAATCCCC CTCAATATCC ATTGACTAGA ATTTAATCAC 18095
 ATGGCCAGCA CAATGCAAG GAAATCTGGA AATATAATC TTTATTCCAG GTAGCCATAT 18155
 GACTCTTTAA AATTGAGAAA TAATATATTT TTAATAATAT ATTCTGGCTT TGGTATAAAG 18215
 AATTGATGCT GTGGGGTGA GAGGGCAAAA TTAAGGGTTG ASAGCCTATT ATTTTAGTTA 18275
 TTACAAGAAA TGATGGTGTG ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT 18335
 GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAAGAC ATGATTTAGG GTTGGATTTG 18395
 GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTTT CACTTAGGCA ATTTGTATCA 19455
 TTCAATGAAA TAGGGAACAC AGGAGGAAGA GCAGGTTTTG GTGTATACAA AGAGGAGGAT 18515
 GSATGAGCCA TTTCGTTTTG GATCTGAGAT GTCTGTGAAA CGTCTAGTG GAGATGTCCA 18575
 CAAACTCTTC TACATGTGCT TCTGAGTTCA GACACAGAT TTGGGCTGGA GATAGAGATA 18635
 TTGTAGGCTT ATACATAGAA ATGGCATTTG AATCTATAGA GATAAAAAGA CACATACAG 18695
 GAAATGTGTA AAGTAGAGTA GAAAGGCTCA ACTACTGTG TGGGGGGAAT ACCTACATTT 18755
 AAAGGATGCA GTAGAAAGAA GCTAATAAAC AACAGAGAGC ASACTAACCA AAAGGAGGAA 18815
 AGAAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG 18875
 TGTGTGTGTT AATTTTGCA CCTTGAGAAAT CAAGGGCCAG AACACAGCTT TTAGATTTAG 18935
 CAACAAGGAG TTTGGTGATC TCAATGAAA GAGCTTGATG GTGAAATGGA GGCAGAGGCA 18995
 GATTGCAATG AGTGAACAG TGAATGAGAA GTGAAAGAA GATACAGATA ATTCTGTCTA 19055
 AAAGCTTGGC TGTAAAGAG AGGAGAGAAA CAAGACTAGC TSCAAAGTGA GATTGGGTTG 19115
 ATGGAGCAGT TTTAAATCTC AAAATAAAGA GCTTTGTGCT TTTTGTATTA TGAAAAATAAT 19175
 GTGTTAATTG TAACATATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAAATAT 19235
 AAAAACCACT CAGAAATAAT GATAGCTACC ATTTGTATAC AATATTTCTA CACTCCTTTC 19295
 TATGTATATA TACAGACACA GAAATGCTTA TATTTTATTT AAAAGGGATT GTACTATACC 19355
 TAAGCTGCTT TTTCTAGTTA GTGATATATA TGAACATCTC TCCATGGCAA CGAGTAATTG 19415
 CAGTTATATT AAGTTTATGA TATTTTCAAA TAAAGGCTTA TCTTTGCTCT TTTTATTTAA 19475
 TCAATTTCTA ATTTGTGAAT GTTTGTTCCT AGTTTGTGTT TGTATTAAAC AATGTTCCCA 19535
 TAAGCATTCG TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTCTTCTA GATATAAAAC 19595
 CAGGAGGTAG AATTGCTGG TTGATAGAA AGAAAGGATG ATTGCCAAAT TAAAGCTTCA 19655
 GTAGAGGGTA CATGCCGAGC ACAATGAGAA TCAGCCCTAG ATACCAGAAA TGGCACTTTC 19715
 TCATTTCCCT TTGAGACAAA AAGGAGAGAG GCAATAAATG TGCTGCCAGA GTTAAATTTG 19775
 TACGTGGAGT AGCAGGAAAT CATTGCTGAA AATGAGAAAC AGAGATGATG TTGTAGAGGT 19835
 CCTGAAGAGA GCAAGAAAAA TTTGAAATGG CCGCTATCAG CTATGAGAA GAGTGTCTAA 19895
 CTGGAAAAA AAGAGAGTAT TGAATTTGG TATGCTTGA ATGGCACCGA TTTGAACGCT 19955
 TGTGGCATTC TTCAACAGCA GCACTCAGCA GGCAGTTTTC GAGTTTTGTA GCAGAAAAC 20015
 AAATAAGTTA GGGATTTAAT ATCTTCCCA AATGAGTAC AAAATGAACT CTGAGATCCA 20075
 GCTGCACAGG GAAGGAGAG AAGACAGGAA GAGGTAGAT AGGAAATACA AGAGTCAGGA 20135
 GACTGGAAGA TGTGTGATA TTAAAGAGAA CATAGAGTTG GAGTAAAGT GTAAGAAAAAC 20195
 TAGAAGGATA AGAGACCGGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCAGA 20255
 GTAGTTCTGA ATGTAACAA GAAATTTAGT GTGCTTTTGA GAGTAGGTTA AAAAACAATA 20315
 GGCAGCTTTA TTGTAGCTAC TTCTGAGACA GAAGATTGTC ATTAATAGTT TTAGAAAACT 20375
 AAAATATATA GCATACCTAT TTGTCAATTA ACAAGAGAAC TATGTATTTT TAAATGAGAT 20435
 TTAATSTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA 20486
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu
 -5 1 5

TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT 20534
 Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile
 10 15 20

GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT 20582
 Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys

25	30	35	
AGA G	GT ATTTTTTTT	ATTGCGCAAA	ATAGAAATGA CTAGCTACTT CTTCCCATTC 20638
Arg Asp			
40			
TSTTTTACTG	CTTACATTGT	TCCGTGCTAG	TCCCAATCCT CAGATGAAAA GTCACAG3AG 20698
TSACAATAAT	TTCATTTACA	GGAAACTTTA	TAAGGCATCC ACGTTTTTTA GTTG333TAA 20758
AAAATTGGAT	ACAATAAGAC	ATTGCTAG33	GTCATGCCTC TCTGAG3CTG CCTTTGAATC 20818
ACCAATCCCT	TTATTGTGAT	TGCATTAACT	GTTTAAAACC TCTATAGTTG GATGCTTAAT 20878
CCCTGCTTGT	TACAGCTGAA	AATGCTGATA	GTTTACCAG3 TGTG3T33CA TCTATCTGTA 20938
ATCCTAGCTA	CTTG33AG33	TCLAGCAG3A	GGATTGCTTG AGGCCAG3AC TTT3AG3CTG 20998
TAGTACACTG	TGATCGTACC	TGTGAATA3C	CACTGCACTC CAG3CT33GT GATATACAGA 21058
CCTTGTCTCT	AAAATTAATA	AAAAAAAAAA	AAAAAACCTT A3GAAA33AA ATT3ATCAAG 21118
TCTACTGTGC	CTTCCAAAAA	ATGAATTCCA	AATATCAAA3 TTA33CTGAG TTGAAGCAGT 21178
GAATGTGCAT	TCTTTAAAAA	TACTGAATAC	TTACCTTAAC ATATATTTTA AATATTTTAT 21238
TTAGCATTTA	AAAGTTAAAA	ACAATCTTTT	AGAATTCATA TCTTTAAAAA ACTCAAAAAA 21298
GTTGCAGCGT	GTGTGTTGTA	ATACACATTA	AACTGTGG33 TTGTTTGTTC GTTTGASATG 21358
CAGTTTCACT	CTGTCAACCA	GGCTGAAGTG	CA3TGCAGTG CAGTGGT3TG ATCTCG3CTC 21418
ACTACAACCT	CCACCTCCCA	CGTCAAGCC3	ATTCTCATGC CTCAGTCTCC CGAGTAG3TG 21478
GGATTACAG3	CAT3CACACC	TTACACCCGG	CTAATTTTTC TATTTTGTAG AGAGCTGGGG 21538
TTTACCACAT	TTG3CCAG3C	T3TCTCAAA	CCCTTAACCT CAAGTGATCT GCCTGCCTCA 21598
GCCTCCCAAA	CAAAACAAAC	ACCCACACAG	TTAATATGT3 TTACAACACA CATGCTGCAA 21658
CTTTTATGAG	TATTTTAATG	ATATAGATTA	TAAAAGGTT3 TTTTAACTT TTAAATGCTG 21718
GGATTACAG3	CAT3AGCCAC	TGTGCCAGGC	CTGAAGTGT3 TTTTAAAAA TGTCTGACCA 21778
GCTGTACATA	GTCTCTTGCA	GACTGSCCAA	GTCTCAAAGT G3GAACAGGT GTATTAAGGA 21838
CTATCCTTTG	GTTAAATTTT	CGCAATGTTT	CCTGTGCAA3 AATCTTCTA ACTAGACTTC 21898
TCATTTATTA	TATTTATTTT	AG AT AAT	GCA CCC CGG ACC ATA TTT ATT ATA 21949
	Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile		
40	45		
AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG	GCT GTA ACT ATC TCT		21997
Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met	Ala Val Thr Ile Ser		
50	55	60	65
GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT	GAG AAC AAA ATT ATT		22045
Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys	Glu Asn Lys Ile Ile		
70	75	80	
TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTTCAAT	CATGTTAATA TAATCAATAT		22103
Ser Phe Lys			
AATTAGAAAT ATAACATTAT TTCTAATGTT AATATAAGTA	ATGTAATTAG AAAACTCAAA		22163
TATCCTCAGA CCAACCTTTT GTCTAGAACA GAAATAACAA	GAAGCAGAGA ACCATTAAAG		22223
TGAATACTTA CTAAAAATTA TCAAACTCTT TACCTATTGT	GATAATGATG GTTTTTCTGA		22283
GCCTGTGCACA GGGGAAGAGG AGATACAACA CTTGTTTTAT	GACCTGCATC TCCTGAACAA		22323
TCAGTCTTTA TACAAATAAT AATGTAGAAT ACATATGTGA	GTTATACATT TAAGAATAAC		22403
ATGTGACTTT CCAGAAT3AG TTCTGCTATG AAGAATGAAG	CTAATTATCC TTCTATATTT		22463
CTACAACCTT GTAAATTATG ATAATATTTT AATCC3TAGT	TGTTTTGTTG CTGATCCTTA		22523
GCCTAACTTT TAGACCAAG CTTCAGCTTC CAGTTGATGT	ATGTTATTTT TAAT3TTAAT		22583
CTAATTGAAT AAAAGTTATG AGATCAG3TG TAAAA3TAAT	GCTATATTTA TCTTCAAGCC		22643
AG3TATAAAG TATTTCT3GC CTCTACTTTT TCTCTATTAT	TCTCCATTAT TATTTCTCTAT		22703
TATTTTTCTC TATTTCCCTCC ATTATTGTTA GATAAAACCA	AATTAACATAT AG3TACABAC		22763
T3AGCCAGTA AGAGTAGCCA G33ATGCTTA CAAATTTGGCA	ATGCTTCAGA GGAGAATTCC		22823
AT3TCAT3AA GACTCTTTTT GAGT33AGAT TT3CCAAATA	ATATCCGCTT TCATGCCAC		22883
CCAGTCCCCA CTGAAAGACA GTTAG3ATAT GACCTTAGTG	AAGGTACCAA G33GCAACTT		22943
G3TAG3BAGA AAAAG3CCAC TCTAAATAT AATCCAAAGTA	AGAACAGTGC ATATG3AACA		23003
GATACAG3CC CCAGACAAAT CCCTCAG3TA TCTCCCTCCA	ACCAGAGTGC CACCCCTTCA		23063
GGTGACAATT TGGAGTCCCC ATTCTAGACC T3ACAG3CAG	CTTAGTTATC AAAATAG3AT		23123
AAGAG3CCTG GGATG3AAG3 GTAG33TGGA AAG33TTAAG	CATGCTGTTA CTGAACAAAC		23183
TAATTAG3AG GGAAG33AT G33CAAG3TC AAG3TATGTT	GGATAGAGGA AAACTCAG3T		23243
G3AGAG3CAG ATTCAAGAAC TG3GATAAGT CC3AACCTAC	AG3TGGATTG TTGTT3A333		23303
AGACT33TGA AAATGTTAAG AAGATG3AAG TAAT3CTTGG	CACTTAGTAG GAAGT333CA		23363
AATCCATATT TGGG33AGCC T3AAGTTTAT TCAATTTTGA	TG3CCCTTTT AAATAAAAA		23423
AATGT33TG G3CGTGGT33 CTCACAC3TG TAATCCAGC	ACTTTG33AG GCG3A3333		23483
GCGGATCACC TGAAGTCA33 AGTTCAAGAC CAGCCTGACC	AACATG33AG AAACCCATCT		23543
CTACTAAAAA TACAAAATTA GTTG33GCT3 GT33CATAT3	CCTGTAATCC CA3CTACTCG		23603
G3AGGCTGAG GCAG3AGAA CTTTTGAACC CG33AG33CAG	AG3T33GAT GA3CCTAGAT		23663
CGTGCCATTG CACTCCAG3C T333CAACAA GAC3CAAACT	CG3TCTCAA AAAAAAAAA		23723
AAAAAGT3AA ATTAACCAAA G3CATTA3CT TAATAATTTA	ATACTGTTTT TAA3TA333C		23783
GGGGGGTG3C TGGAAGAGAT CT3TGTAAT	GAG3GAATCT GACATTTAAG		23843
ATCATAG3AA ATCTGCTTCT G3AAGGAAT CAATAAATAT	TAGTTG33AG GGG33A3A3A		23903
GTGAGG33TG GACTAGGACC AGTTTTAG3C CTTGTCTTTA	ATCCCTTTTC CT3CCACTAA		23963

TAAGGATCTT	AGCAGTGGTT	ATAAAAAGTGG	CCTAGGTTCT	AGATAATAAG	ATACAAACAGG	24023
CCAGGACACG	TGGCTCATGC	CTATAATCCC	AGCACTTTGG	GAGGGCAAGG	CGAGTGTCTC	24083
ACTTGAGATC	AGGAGTTCAA	GACCAGCCTG	GCCAGCATGG	CGATACTCTG	TCTCTACTAA	24143
AAAAAATACA	AAAATTAGCC	AGGCATGGTG	GCATGCACCT	GTAATCCCAG	CTACTCGTGA	24203
GCCTGAGGCA	GAAGAATCCG	TTGAAACCCG	GAGGTGTAAG	CTGCAGTGAG	CTGAGATCGC	24263
ACCACTGGAC	TCCAGCCTGG	GCGACAGAAAT	GAGACTTTGT	CTCAAAAAAA	GAAAAAGATA	24323
CAACAAGCTA	CCCTTATGTG	CTCACCTTTT	ACTGTTGATT	ACTAGCTATA	AAGTCCTATA	24383
AAGTTCTTTG	GTCAGAAGCC	TTGACAAACAC	TAAGAAGGGAT	TTGCTTTGAG	AGGTTACTGT	24443
CAGAATCTGT	TTCATATATA	TACATATACA	TGTATATATG	TATCTATATC	CAGGCTTTGG	24503
CAGGCTTCCC	TCACTACTTT	CAGTGCACCT	GGAGATGTTT	AGGTCAATAT	CAACTTTCCC	24563
TGSATTCAGA	TTCAACCCCT	TCTGATGTAA	AAAAAAGAAA	AAAAAAGAAA	GAAATCCCTT	24623
TCCCTTTGGA	GCACTCAAGT	TTCAACAGGT	GCGGCTTTCC	AAGTTGGGGG	TTCTCCAAAG	24683
TCATTGAGAT	TGCTTTTACA	TCCATTTGCT	ATGTACCTTC	CCTATGATGG	CTGAGAGTGG	24743
TCAACATCAA	AAGTAGGAAA	GCTACTGCCC	AAGGATGTCC	TTACTCTAT	TCTGAAATGT	24803
GCAATAAGTG	TGATTAAAGA	GATTGCTGTG	TCTACCTATC	CACACTCTCG	CTTTCAACTG	24863
TAACCTTTCT	TTTTTCTTTT	TTTTTCTTTT	TCTTTTCTTT	TGAAAGGGAG	TCTCGCTCTG	24923
TGCGCCAGGC	TAGAGTGCAG	TGGCACGATC	TCAGCTCACT	GCAAGCTCTG	CCTCCCGGCT	24983
TCAGGCTCATT	CTCCTGCTCT	AGCCTGCCAA	GCAGCTGGGA	CTACAGGCGC	CTGCAACCAT	25043
GCCCAAGTAA	TTTTTTGTAT	TTTTAGTAGA	GACGGGCTTT	CACCGTGTTA	GCCAGGATCG	25103
TCTCGATCTC	CTGAACCTGT	GATCGGCTCG	CCTCAGCTCT	CCAAAGTGCT	GGGATTACAG	25163
GCGTGAGCCA	TCCGACCCCG	CTCAACTGTA	ACTTTCTATA	CTGCTTCATC	TTCCCTGTGA	25223
ATGTTACTAG	AGCTTTTGAA	GTTTTGGCTA	TGGATTATTT	CTCATTATTA	CATTAGATTT	25283
CAGATTAGTT	CCAAATTGAT	GCCACAGCTG	TAGGCTCTCT	TCTTAAATTG	TATATTGTAG	25343
ACAGCTGCAG	AAGTGGGTGC	CAATAGGGGA	ACTAGTTTAT	ACTTTTCATC	ACTTAGGACC	25403
CACACTTGTT	GATAAAGAAG	AAAGGTCAAG	AGTTATGACT	ACTGATTCCA	CAACTGATTG	25463
AGAAGTTGGA	GATAACCCCG	TGACCTCTGC	CATCCAGAGT	CTTTCAAGCA	TCTTTGAAAG	25523
ATGAAGAAAT	GCTATTTTAA	TTTTGGAGGT	TTCTCTATCA	GTGCTTAGGA	TCATGGAAT	25583
CTGTGCTGCC	ATGAGGCCAA	AATTAAGTCC	AAAACATCTA	CTGCTTCCAG	GATTAAACAT	25643
GAAAGAACCT	AAGTGGTGCC	CACATGTTCT	GATCCATCCT	GCAAAATAGA	CATGCTGCAC	25703
TAACAAGAAA	AAGTGACGCA	GCACTACCA	TTGATAAACC	TGCAAGATTA	TAGTTTCAAG	25763
TAATCTAACC	ATTCTTACA	AGGCCCTATT	CTGTGACTGA	AACATACAAG	AATCTGCTAT	25823
TGGCTTTCTA	AGGCAGGGCC	CAGCCAAGGA	GACCATATTC	AAGACAGAAA	TTCAAGACTA	25883
CTATGAAACT	GAGTGCTTGG	GCAAGGAAAG	CAGAGTCAA	GACTGCAAC	TGAGCCAAAT	25943
CAGAGGCTT	ACACAGGAA	CCAGGCTCTA	GCCCTACAAC	AATTATTGGG	TCTATTCACT	26003
GTAAGTTTAA	ATTTCAAGCT	CCACTGAAAG	AGTAAGCTAA	GATTCTGTGG	ACTTTCTGTC	26063
TCTCTCAGAG	TTGGCTCAGA	AATGAGAACT	GCTCAGGCGA	GCCATGCTGG	CTTACAGCTG	26123
GAATCCCAGC	ACTTTGGGAG	GCCGAAGTGG	GAGGCTCACT	TGAGGCGAGG	AGTTCAAGAC	26183
CAGCTTAAGC	AACAAAGTGA	GATACCCCTT	GACCCCTTCT	CTACAAAAAT	AAATTTTAAA	26243
AATTAGCCAA	ATGTGGTGGT	GTATACTTAC	AGTCCACAGT	ACTCAGGAGG	CTGAGGCGAG	26303
GGGATTGCTT	GAGCCAGGGA	ATTCAAGGCT	GCAATGAGCT	ATGATTTTAC	CAGTGTACTT	26363
CTGCTGAGG	AACAGAGCGA	GACCCGTGCT	CATAGCAAAA	AGAAAAAGAA	ACTAGACTTA	26423
GCCTAAGTTT	GTGGGAGGAG	GTCATCATCG	TCTTTAGCGG	TGAATGGTTA	TTATAGAGGA	26483
CAGAAATTGA	CATTAGCCCA	AAAAGCTTGT	GCTCTTTGCT	GCAACTCTAC	TTAATCTTGA	26543
GCAAAATGTG	ACACCACTCA	ATGGAAGAGG	AGAGAAATTA	GCTGTTTGAT	GTATAGGGGA	26603
AAACTAAGG	CCTGGAAGTG	AATATGCTAT	CCATGAGAGG	GAGAAATAGG	GATTGCGAGT	26663
TAAGAAAGAG	AGGAGCTGAG	TACTGCTGTT	CAGAGATTTT	TTTTATGTAA	CTCTTGAAGG	26723
GCAAAACTAC	TTTTGTTCTG	TTTGTAATA	TACTTCAAAA	CAAACTTCAT	ATATTCAAA	26783
TGTTTATGTC	CTGAAATAAT	TAGGTAATGT	TTTTTTCTCT	ATAG GAA ATG AAT CTT		26839

Glu Met Asn Pro

85

CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA	26887
Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg	
90 95 100	
AGT GTT CCA GGA CAT GAT AAT AAG ATG CAA TTT AAA TCT TCA TCA TAC	26935
Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr	
105 110 115 120	
GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC	26983
Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu	
125 130 135	
ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT	27031
Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr	
140 145 150	
GTT CAA AAC GAA GAC T AGCTATTAAA ATTTTCATGCC GGGCGAGTG GCTCAGCCT	27087
Val Gln Asn Glu Asp	
155	
GTAATCCCAG CCCTTTGGGA GGGTGAGGCG GGCAGATCAC CAGAGGTCAG GTGTTCAAGA	27147
CCAGCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTGAGT	27207

GTAGTGACCC	ATGCCCTCAA	TCCCAGCTAC	TCAAGAGGCT	GAGGCAGGAG	AATCACTTGC	27267
ACTCCG3AGG	TGGAGGTTGT	GGTGAGCCGA	GATTGCACCA	TTGCGCTCTA	GCCTGGGCAA	27327
CAACAG3AAA	ACTCCATCTC	AAAAAATAAA	ATAAATAAAT	AAACAAATAA	AAAATTCATA	27387
ATGTGAACTG	TCTGAATTTT	TATGTTTGA	AAGATTATGA	GATTATTAGT	CTATAATTGT	27447
AATGGT3AAA	TAAAATAAAT	ACCACTCTTG	AAAAACATCA	TTAAGAAATG	AATGAACTTT	27507
CACAAAAGCA	AACAAACAGA	CTTTCCCTTA	TTTAAGTGAA	TAAAATAAAA	TAAAATAAAA	27567
TAATGTTTAA	AAAATTCATA	GTTTGAAAAC	ATTCTACATT	GTAAATTGGC	ATATTAATTA	27627
TACTTAATAT	AATTATTTTT	AAATCTTTTG	GGTTATTAGT	CCTAATGACA	AAAGATATTG	27687
ATATTTGAAC	TTTCTAATTT	TTAAGAATAT	CGTTAAACCA	TCAATATTTT	TATAAGGAGG	27747
CCACTTCACT	TGACAAATTT	CTGAATTTCC	TCCAAAGTCA	GTATATTTTT	AAAATTCAGT	27807
TTGATCCTGA	ATCCAGCAAT	ATATAAAGG	GATTATATAC	TCTGGCCAAC	TGACATTCAT	27867
CCTAGGAATG	CAAAGATGGT	TTAATATCCT	AAAATCAATT	AACATAACAT	ACTATATTAA	27927
TAAAGTATCA	AAACAGTATT	CTCATCTTTT	TTTCTTTTTT	CACAATTCCT	TGGTTACACT	27987
ATCATCTCAA	TAGATGCAGA	AAAAGCATTT	GACAAAATCC	AATTCATAAT	AAAAATTCTC	28047
AAACTT3AAA	GAGAACATCA	TAAAG3CATC	TATGAAAAAC	CTACAGCTAA	TATCATACTT	28107
AACGATGAAA	AACTGAATTA	TTTTACCCTA	AGATCAAGAA	TAATGCAAGC	ATGTCAGCTC	28167
TTGCAACTTC	TATTCAACAT	TGTACTGGAG	GTTCTAGCCA	GAGCAACCAT	ACAATAAATA	28227
AAAATAAAAG	GCA3CCAGAT	TAGAAAGGAA	GTCTTTATTT	GCAGACAACA	TGGTTCTTTA	28287
TGCAGAAAC	CGTCAGGAAT	ACACACACAT	GTTAGAACTA	ATAAGTTCAG	CAAGGTTGCA	28347
GGTTGCAAT	TCAATATGCA	AAAATACATT	GAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA	28407
CCTTTCTGTC	CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTTG	28467
AGGCTATAGT	GCAATGTGAT	CTTGCTGTG	AATAGCCACT	GCACTCGAGC	CTAGGCAACA	28527
AAGTGAGACC	CCGTCTCCAA	AAAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAT	28587
GAATGATCTG	AAAACAAGAA	AATTCCATTC	ACGATGGTAT	TAAAAAATA	AAATACAAAT	28647
AAATTAGCA	AAATAATTAT	AAAACCTGTA	CATCGAAAAAT	TTCAAAGCAC	TCTGAGGGAA	28707
ATTAAAGATG	ATCTAAATAA	TTGGAGAGAC	ACTCTATGAT	CACTGATTGG	AAAATTCATT	28767
CAATATTGTT	AAGATAACAA	TTGTCCCAAA	ATTGATGCAT	GCATTCAATT	TAGTCTTCAT	28827
CAAAATTCCA	GCAGGGTTTT	TG3AGAAATT	GACAAGCTGT	ACCCAAAATG	TATATGGAAA	28887
TGAAAAGACC	CAGAAGAGCA	AATAATTTTT	TAAAAACAAA	GTTGGAAAAC	TTTACTTCC	28947
TAATTTTAAA	ACTTACTATA	AACTAAAGT	TATCAAGACC	ATTTAGT		28994

(15) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminal fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Tyr	Phe	Gly	Lys	Leu	Glu	Ser	Lys	Leu	Ser
1				5				10	

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCATCCTAAT ACGACTCACT ATAGGCG

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATAGGGCA CGCGTGGT

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAGTTTTC ACCTTCCAAC TGTAGAGTCC

30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGATCAAGT CGTGATCAGA AGCAGCACAC

30

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTGGCTGCC AACTCTGGCT GCTAAAGCGG

30

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTATTGTCAA TAAATTCAT TGCCACAAAG TTG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGATGGCTG CTGAACCAAGT AGAAGACAAT TGC

33

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCTTGGTCA ATGAAGAGAA CTTG3TC

27

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC

33

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCAGCCTAG AGGTATGGCT GTAACATCT C

31

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATGAAAT TTTAATAGCT AGTCTTCGTT TTG

33

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGACATCAT ATTCTTTCAG AGAAGTGTC

30

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAATTTGAA TCTTCATCAT ACGAAGGATA C

31

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCCGAAGCTT AAGATGGCTG CTGAACCACT A

31

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAAATAATT TTGTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTAGCGGC CGCGGCATGA AATTTTAATA GCTAGTC

37

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC

33